

DATU-BASE BANATUAK

Bibliografia

- Elmasri, R. & Navathe, S.B.

Datu-base sistemak. Oinarriak

Euskal Herriko Unibertsitateko Argitalpen Zerbitzua (2009)

- Jatorrizko bertsioa – ingelesez; gazteleraz ere badago.

- Gehiago sakondu nahi izanez gero

- Valduriez, O.

Principles of Distributed Database Systems

Prentice Hall International Inc. (2008)

- García Molina, H., Ullman, J.D., Widom, J.

Database Systems. The Complete Book.

Pearson International (2009)

1. Motibazioa
2. DBBen ezaugarriak
3. DBBen helburua
4. DBBen diseinua

1. DB banatuak. Motibazioa.

- DBBen teknologia, bi teknologien baterakuntzatik sortu zen (itxuraz kontrajarriak):
 - 1) Datu-baseen teknologia (tradizionala)
 - Datuak zentralizatuta (toki ezberdinetatik atzitu baina toki bakarrean kokatuta)
 - Hainbat fitxategi → Datu-base bat
 - 2) Sareko eta datu-komunikazioko teknologia
 - Banaketa/baliabideen partekatzea
 - DB zentralizatua → DB banatua

DB teknologiak datuen INTEGRAZIOA bilatzen du eta ez ZENTRALIZAZIOA

DB banatuak vs. zentralizatuak

- Datu-base sistema **zentralizatuen** eta **banatuen** arteko diferentzia nagusia, honetan datza: zentralizatueta datuak **kokaleku bakarrean** daude eta **banatueta hainbat kokalekuean**.

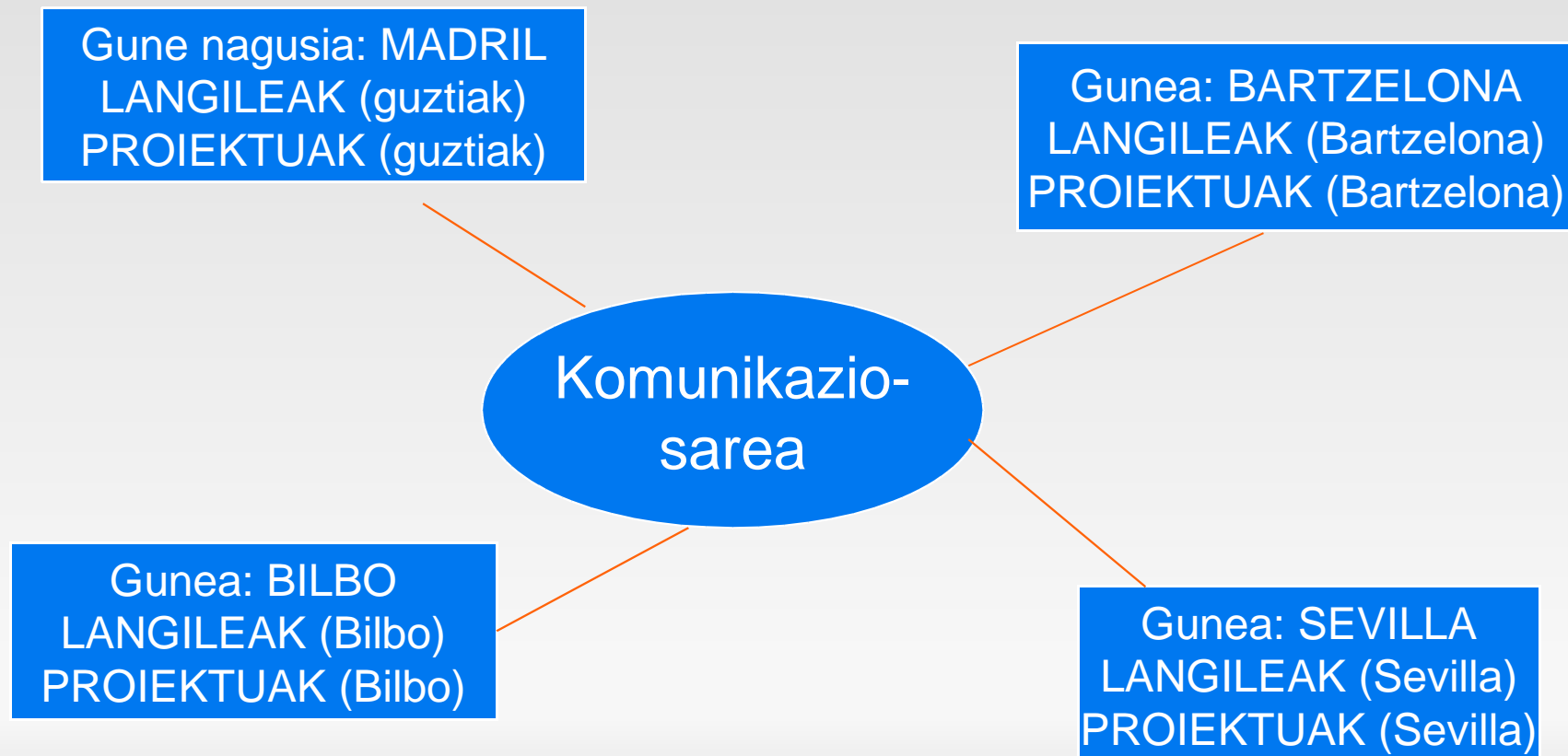
Datuen banatze honek kontsulten eta transakzioen prozesamenduan zailtasun anitz sortzen ditu.

2. DB banatuen ezaugarriak. Definizioa

- 3 definizio posible:
 - Logikoki sistema berekoak diren datuen bilduma, fisikoki sare bateko leku ezberdinetan banatuta daudelarik.
 - DBB bat konputagailu-sare batean banatuta eta elkarrekin logikoki lotuta dauden hainbat datu-baseren bilduma da.
 - Posizio fisiko bakarrean biltegitratuta **ez** dagoen DBa. Geografikoki banatuta dauden leku fisikoetan banatuta dago eta komunikazio-loturen bidez konektatuta, **sare-adabegi** bakoitzeko erabiltzaileek datu guztiak atzi ditzaketelarik.

DB banatuak. Aukera posible bat. Adibidea (Datuen zatikatzeta eta esleipena)

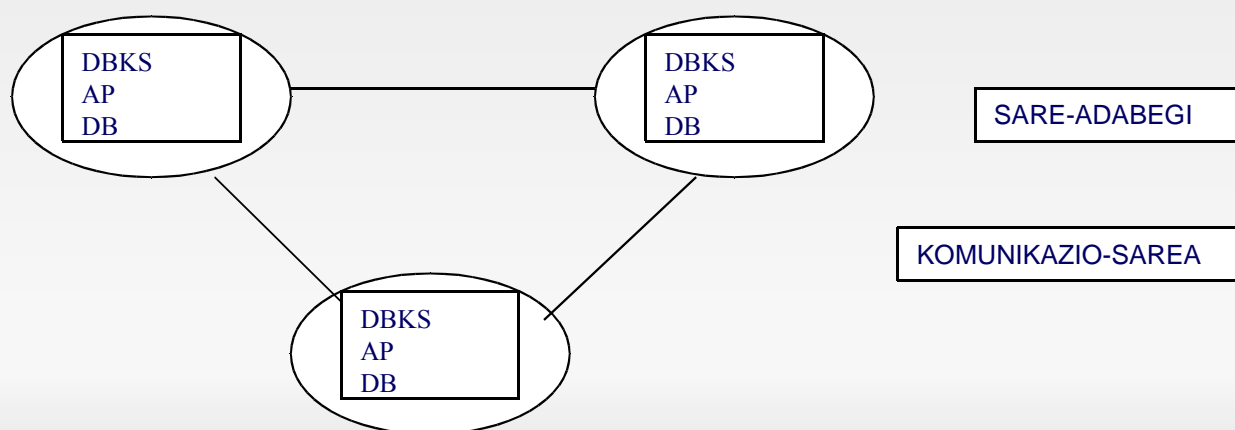
Guneetako batean datu-base zentralizatua duen sare-arkitektura. LANGILEAK eta PROIEKTUAK taulak horizontalki zatikatuak eta errepikatze posibleekin biltegitatuak.



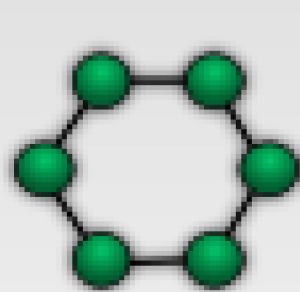
- **Datu-base banatuak kudeatzeko sistema (DBBKS)** bat datu-base banatu bat kudeatzen duen eta, aldi berean, banaketa erabiltzailearentzako garden bihurtzen duen software-sistema da.
- **Gardentasunak** zera esan nahi du: aplikazioek, alderdi logikoari dagokionez, makina bakarrean exekutatutako DBKS bakar batek datu horiek kudeatuko bailituen lan egingo dute.

DBBen sistema baten arkitektura

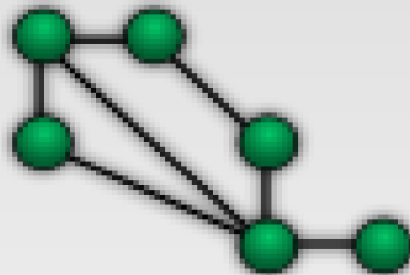
- Sare-adabegi multzo bat da. Bakoitzak bere DBS lokala du eta hainbat sare-adabegietako datuak atzitzen dituzten kontsulta/transakzioen exekuzioan parte har dezakete.
- Aplikazio-prozesatzailea (AP). Banaketa-lanak egiten dituen softwarea.
- DBBS = DBB + DBBKS



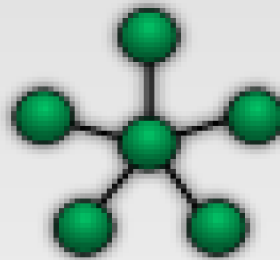
Konfigurazioa sarean



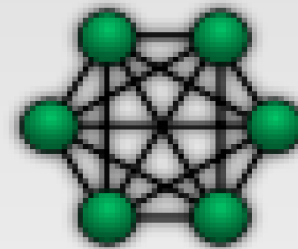
Ring



Me sh



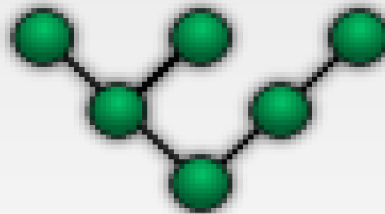
Sta r



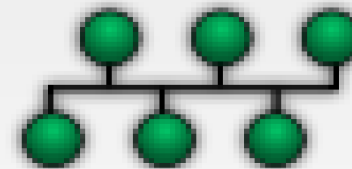
Fully Connected



Line



Tree



Bus

Konfigurazioa sarean

- Konfigurazioak haien artean konpara daitezke, ondorengo irizpideak kontuan hartuta:
 - **Instalazio-kostua**: sistemako adabegiak fisikoki lotzeko kostua
 - **Kostua komunikazioetan**: A sare-adabegitik B sare-adabegira mezu bat bidaltzeko kostua denboran eta dirutan
 - **Eskuragarritasuna**: nahiz eta loturaren edo sare-adabegiren batean akatsak egon, datuak atzitzeko dugun gaitasuna

Arkitekturako sare-adabegiak

- Sare-adabegietan honakoak egon daitezke:
 - **Erabiltzaile/transakzio lokalak**. Beraien sare-adabegitik soilik atzitzen dituzte datuak.
 - **Erabiltzaile/transakzio globalak**. Sare-adabegi ezberdinetan biltegiatutako datuak atzitzeko beharra dute.

3. DB banatuen helburua

- **Gardentasuna:** Sistemaren maila semantikoaren (maila altua) eta inplementazio-kontuen (maila baxua) banaketa

DATUEN INDEPENDENTZIA eskaintzea

- Gardentasun **motak**:
 - Banaketaren edo sarearen gardentasuna
 - Errepikatzearekiko gardentasuna
 - Zatikatzearekiko gardentasuna

Banaketaren gardentasuna

- Erabiltzaileak ez du sarearen funtzionamenduari buruzko xehetasunik ezagutu behar. Honakoetan banatzen da:
 - ***Kokalekuarekiko gardentasuna:*** Ataza bat egiteko agindua, independentea da datuen kokalekuarekiko eta agindua jaulki den sistemaren kokalekuarekiko.
 - ***Izendatzearekiko gardentasuna.*** Objektuei izenak ematen zaizkienean, objektu izendunak anbiguotasunik gabe atzi daitezke.
- **Kokatze-eskema** batek datuen (zatien) kokalekua deskribatzen du.

Errepikatzearekiko gardentasuna

- Adibidean ikusi dugun moduan, datuen kopiak hainbat lekutan gorde daitezke, atzigarritasuna, errendimendua eta fidagarritasuna hobetzeko. Beharrezko datua atzitzeko denbora murrizten da.
- Erabiltzaileak ez du jakin behar kopiarik baden edo ez. Kopiak topatu eta mantentzeko xehetasun guztiak sistemak landu behar ditu.
 - Abantailak: Segurtasuna (*eskuragarritasuna*), komunikazio kopuru txikiagoa (*paralelismoa handitzea*).
 - Desabantailak: Aldatzeak, biltegi-espazio handiagoa.

Errepikatzea eta zatitzea

- **Errepikatzea**: Sistemak erlazio baten kopiak (errepikatzeak) gordetzen ditu, kopia bakoitza toki ezberdin batean gordetzen duelarik. Kontrajarriaz, erlazioaren kopia bakarra ere gorde daiteke.
- **Zatikatzea**: Sistemak erlazioa zati ezberdinetan gordetzen du eta zati bakoitza toki ezberdin batean gordetzen du.
- **Errepikatzea eta zatikatzea konbinatu ohi dira**: Erlazioak zatitan banatzen dira eta zati bakoitza errepika daiteke.

Datuen kopiak

- Aukera ezberdinak:
 - a) DB osoa sare-adabegietako bakoitzean kopiatzea (**errepikatzea**). DBB guztiz errepikatua sortzen da.
 - Abantailak:
 - Sare-adabegietakoren bat badabil, sistema badabil.
 - Paralelismoa handitzen da.
 - Kontsulta orokorretarako, datuak jasotzea azkarragoa da.
 - Desabantailak:
 - Aldaketek kostu handiagoak.
 - Konkurrentziaren kontrolerako eta akatsen aurrean berreskuratzeko mekanismo konplexuagoak.

Datuen kopiak

- b) Datuen originalak edukitzea, inolako kopiarik ez. Zati bakoitza toki bakarrean dago. Banaketa ez erredundantea.

Datuen kopiak

c) Errepikatze partzialak. Zati batzuk errepikatzen dira eta beste batzuk ez.

Sare-adabegi batean, bitan... edo sare-adabegi guztietan errepikatzen dira.

Errepikatutako zatien deskribapenari, **errepikatze-eskema** deitzen zaio.

Tokien aukeraketa eta errepikatze maila honakoen arabera da: sistemaren errendimenduari eta erabilgarritasunari ipini dizkiogun helburuen arabera da, baita sare-adabegietako bakoitzera bidalitako kontsulten mota eta maiztasunaren arabera ere.

Zatikatzearekiko gardentasuna

- Bi zatikatze mota:
 - **Zatikatze horizontala:** erlazio bat tupla (errenkada) multzotan banatzen du.
 - **Zatikatze bertikala:** erlazio bat azpierlazioetan banatzen da, non azpierlazio bakoitza jatorrizko erlazioko zutabeen azpimultzo batek definitzen baitu.
- Gardentasuna: erabiltzaileak kontsulta global bat egitean, kontsulta hori hainbat kontsulta zatitan zatikatu behar da eta erabiltzaileak ez du horretaz konturatu behar. Ez du zatirik baden ezagutu behar.

Helburuak. Laburpena.

Banaketaren gardentasuna

+

Errepikatzearekiko gardentasuna

+

Zatikatzearekiko gardentasuna

Erabiltzailearentzako sistema banatuak,
sistema zentralizatu moduan funtziona dezan
ahalbidetzen du.

Helburuak. Laburpena.

Gainera,

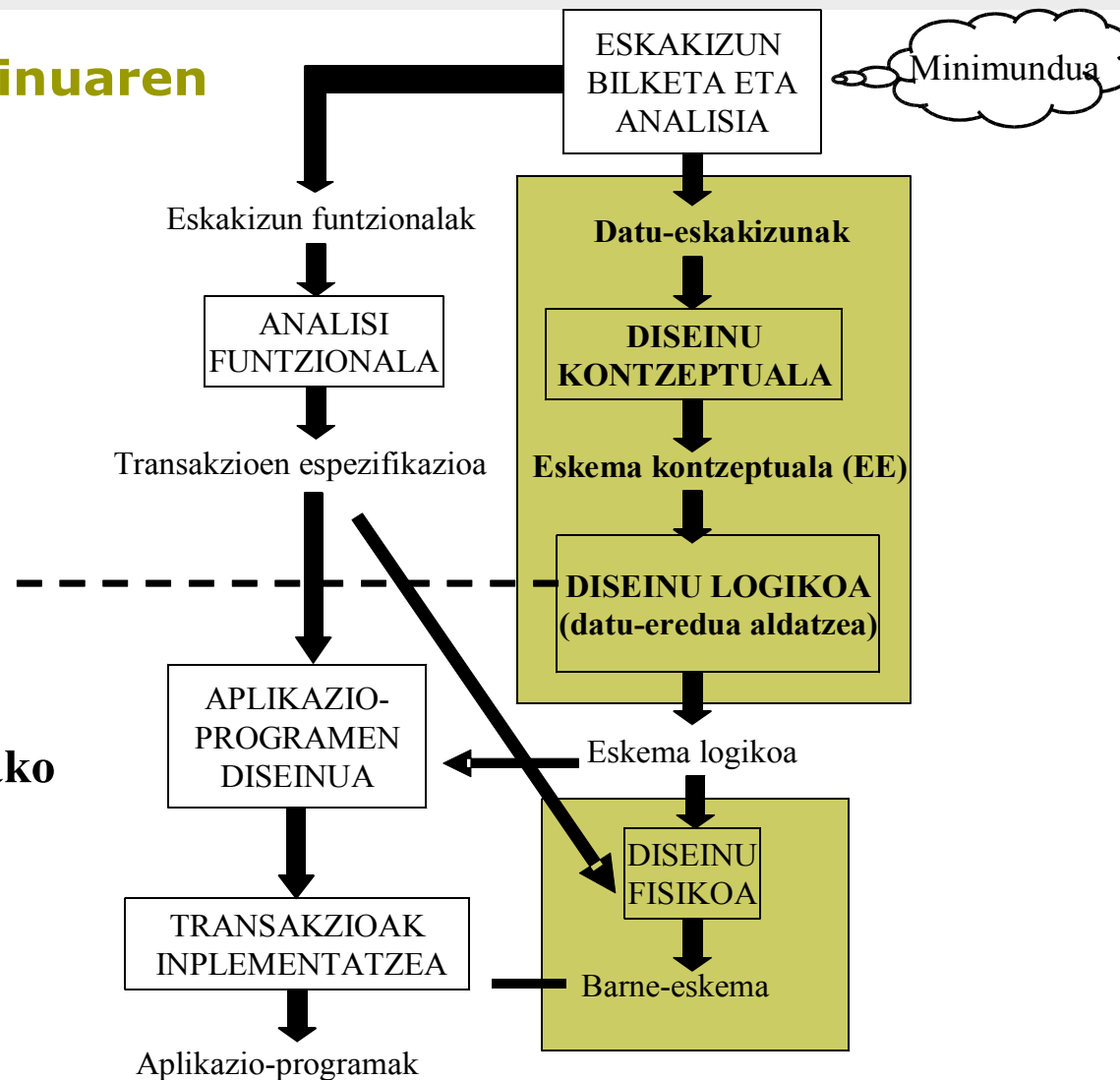
- Kokalekuarekiko gardentasunari esker, erlazioak edo taulak sare-adabegi batetik bestera mugi daitezke, aplikazioak aldatu gabe.
- Kopiekin dagoen gardentasunak datuen eskuragarritasuna handitzen du baita sistemaren errendimendua hobetu ere.
- Zatikatzearekiko gardentasunak ere sistemaren errendimendua hobetzen du.

4. DB Ben diseinua. Memoria egiten.

DB baten diseinuaren faseak

DBKSrekiko independentea

DBKS bakoitzarako zehatza



4. DBBen diseinua

- Honakoak zein sare-nodotan kokatu erabaki behar da:
 - 1) Datuak — **DBBen diseinua**
Eskema kontzeptual orokorra lortu ondoren, **zatikatu** eta **esleitu** egin behar da.
 - 2) Datuekin dabilzan aplikazioak
- DB banatuen kudeaketarako sistema bat behar da, honakoa egingo duena:
 - Kotsultak prozesatzea, kontsistentzia mantenduz, datuen errepikatzea, transakzioen kontrola e.a. badaude.

DBBen diseinua.

Goitik beherako (top-down)

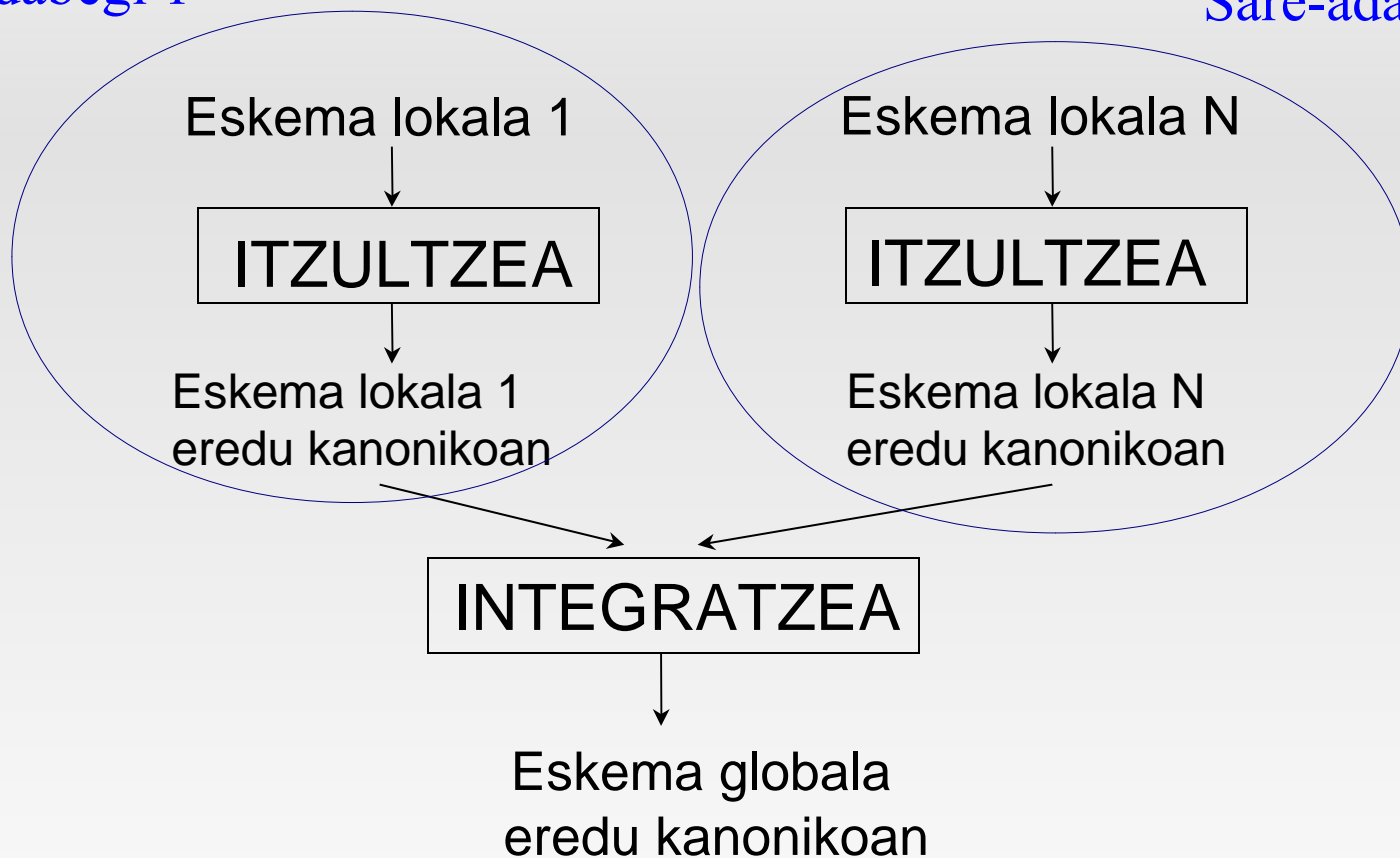


DBBen diseinua.

Behetik gorako (Bottom-up)

Sare-adabegi 1

Sare-adabegi N



Goitik beherako (top-down) DBBen diseinua. Etapak.

1. Eskema globalaren diseinua
2. Zatikatzea eta esleitzea
3. Diseinu fisikoak

Goitik beherako (top-down). DBBen diseinua

- Eskema globaletik abiatuta, eskema lokalak lortzearen arazoak bi zati ditu:
 - **Zatikatzea**: erlazio edo taulak zatitan banatzea.
 - **Esleitzea**: zatiak eskema lokalen artean banatzea.
- “Zatia” banatu beharreko unitatea da eta finkatu egin behar dira.
 - Abantaila: transakzioen konkurrentzia maila handitzen du.
 - Desabantaila: transakzio batzuk degradatu egingo dira zati batzurekin lan egin behar badute.

Zatikatzea

- R erlazio baten zatikatzea **zuzena** da, honako propietateak betetzen badira:
 - **Osotasuna**: R-ko elementu guztiek, R_i -ko zatietakoren batean agertu behar dute.
 - **Leheneratzea**: R_i -ko zatietatik abiatuta, Rren edukia lortu behar da.
- Zatikatze egoki bat egiteko, honako informazioarekin egin behar da lan:
 - Datuen esanahiarekin
 - Datuak erabiltzen dituzten aplikazioekin
 - Komunikazio-sarearekin

Zergatik zatitu?

- **Abantailak:**
 - Aplikazioen errendimendua hobetzen da, erlazioen azpimultzoekin lan egiten baita.
 - Zatiei esker, aldi bereko exekuzio kopurua handitzen da.
- **Desabantailak:**
 - Zati batzurekin lan egiten duten aplikazioen eraginkortasuna txikitzen da.
 - Integritate-murriztapenak egiaztatzea garestiagoa izan daiteke.

Zatikatzea

Zatikatze horizontala: hautespen baldintzak topatzen oinarritzen da



Zatikatze bertikala: Proiektatu ahal izateko atributu multzoak topatzean oinarritzen da.



Zatikatze mistoa (hibridoa): Aurreko bien nahasketan oinarritzen da.

Zatikatze horizontala

- **Zatikatze horizontalak** erlazio bat zatikatzen du errenkadak taldekatu eta azpimultzoak sortuz, non azpimultzo bakoitzak esanahi logiko bat baitu.
- Erlazio bateko **zati horizontala** erlazio horretako tuplen azpimultzo bat da. Zati horizontalari dagozkion tuplak erlazioko atributu batekin edo batzurekin definitutako baldintza baten bidez zehazten dira.
- Eragilea: Hautespena (δ).
- Adibidea: LANGILEA taulan 3 zati horizontal defini ditzakegu baldintza hauekin: (SZ = 5), (SZ = 4) eta (SZ = 1).

Zatikatze horizontala

- Propietateak:
 - Osoa. R erlazio originaleko tupla bakoitza zatietakoren batean dago (C_1 or C_2 or C_n)
 - Disjuntua. Askotan, zatikatze horizontal osoak disjuntuak ere badira. R-ko tupla bat ez dago bi azpimultzotan. Edo, edozein $i \neq j$ -rako, inolako tuplarik ez dago C_i -n eta C_j -n.

Zatikatzea osoa eta disjuntua bada, UNION eragiketa erabil daiteke hasierako erlazioa (originala) berreskuratzeko.

Zatikatze bertikala

- **Zatikatze bertikalak** erlazio bat zutabeetan zatikatzen du. Erlazio bateko **zati bertikal** batek erlazioko atributu jakin batzuk ditu (atributu hauek nolabaiteko erlazioa dute haien artean).
- Eragilea: Proiektzioa (π).
- Adibidea: LANGILEA taulan 2 zati bertikal:
 - Informazio pertsonala: NA, Izena, Jaioteg, Helbidea eta sexua.
 - Lanari lotutako informazioa: NA, Lansaria, GainNA, Szk.
- Beharrezkoa da zati bertikal **guztietan oinarrizko gakoa** edo gako hautagairen baten atributuak sartzea, 34 erlazio osoa berreraiki ahal izateko.

Zatikatze bertikala

- Zatikatze bertikala **osoa** da honakoa betetzen bada:
Zati bertikalen multzoak erlazio originaleko atributu guztiak ditu eta zati guztiek oinarrizko gakoa baino ez dute partekatzen.

$L_1 \cup L_2 \dots \cup L_n = \text{ATRIB}(R)$, hau da, R-ko atributu guztiak eta

$L_i \cap L_j = \text{OG}(R)$, edozein $i \neq j$ izanik

- Hasierako erlazioa (originala) berreskuratzeko KANPO-BILKETA eragiketa erabili behar da.

Zatikatze mistoa (hibridoa)

- Bi zatikatze motak elkarrekin nahasiz lortzen da.
- Orokorrean, erlazio bateko zati bat SELECT-PROJECT eragiketen konbinazioaz zehatz daiteke.

Baldin $B = \text{true}$ and $L \neq \text{ATRIB}(R) \Rightarrow$ Zati bertikala

Baldin $B \neq \text{true}$ and $L = \text{ATRIB}(R) \Rightarrow$ Zati horizontala

Baldin $B \neq \text{true}$ and $L \neq \text{ATRIB}(R) \Rightarrow$ Zati mistoa

Baldin $B = \text{true}$ and $L = \text{ATRIB}(R) \Rightarrow$ Erlazio originala

B = Baldintza ($B = \text{true}$, tupla guztiak)

$\text{ATRIB}(R)$ = R -ko atributuak ($L = \text{ATRIB}(R)$, atributu guztiak)

Zatikatze-eskema

- DB baten **zatikatze-eskema** bat datu-baseko atributu eta tupla guztiak biltzen dituen zati multzo baten definizio bat da, eta baldintza hau betetzen du: datu-base osoa berreraiki daiteke zatietatik abiatuz, KANPO-BILKETA eta BILKETA eragiketen sekuentziaren bat aplikatuz.

- Nola esleitu zatiak sare-adabegietara? Datuen errepikatzearen eta kokalekuaren aukeraketa ondorengoen baitan egongo da:
 - sistematik lortu nahi den errendimenduaren baitan
 - lortu nahi den datuen eskuragarritasun mailaren baitan eta,
 - sare-adabegi bakoitzeko transakzioen motaren eta maiztasunaren baitan.

- Esleitze-fasean ondorengoei buruzko informazio kuantitatiboa ezagutu behar da:
 - DBa,
 - erabiliko diren aplikazioak,
 - komunikazio-sarea,
 - sareko sare-adabegi bakoitzak prozesatzeko eta biltegiratzeko dituen gaitasunak.

Esleitzea

- Errepikatzea erabilgarria da datuen atzigarritasuna hobetzeko
- Zatiak eskema lokaletara honela eslei daitezke:
 - Errepikatzerik gabe: zati bakoitza sare-adabegi bakarrean gordetzen da
eguneratzeetarako ona, kontsultetarako txarra
 - Erabateko errepikatzearekin: zati guztiak sare-adabegi guztietan gordetzen dira
konsultetarako ona, eguneratzeetarako txarra
 - Errepikatze partzialarekin: zati batzuk sare-adabegi batean baino gehiagoetan gorde daitezke
konsulten eta eguneratzeen arteko konpromezua

**Kontsultak baino eguneratze gehiago badaude,
orduan errepikatzeak abantaila gutxiago izango ditu**

Esleitzea

	ERABATEKO ERREPIKATZEA	ERREPIKATZE PARTZIALA	ERREPIKATZERIK EZ
KONTSULTAK PROZESATZEA	Errazagoa	Zailagoa	Zailagoa
KONKURRENTZIAREN KONTROLA	Zaila	Zailagoa	Errazagoa
DATUEN ATZIGARRITASUNA	Oso handia	Handia	Baxua
			41

4. Diseinua. Katalogo orokorra

- Non gorde katalogo orokorra?
 - a) Zentralizatua. Sare-adabegi bakarrean
 - b) Guztiz errepikatua. Adabegi bakoitzean
 - c) Banatua. Sare-adabegi bakoitzean, adabegirako beharrezkoa den informazioa gordetzen da
- a eta c-ren konbinazioa.