

# **DATU-BILTEGIAK**

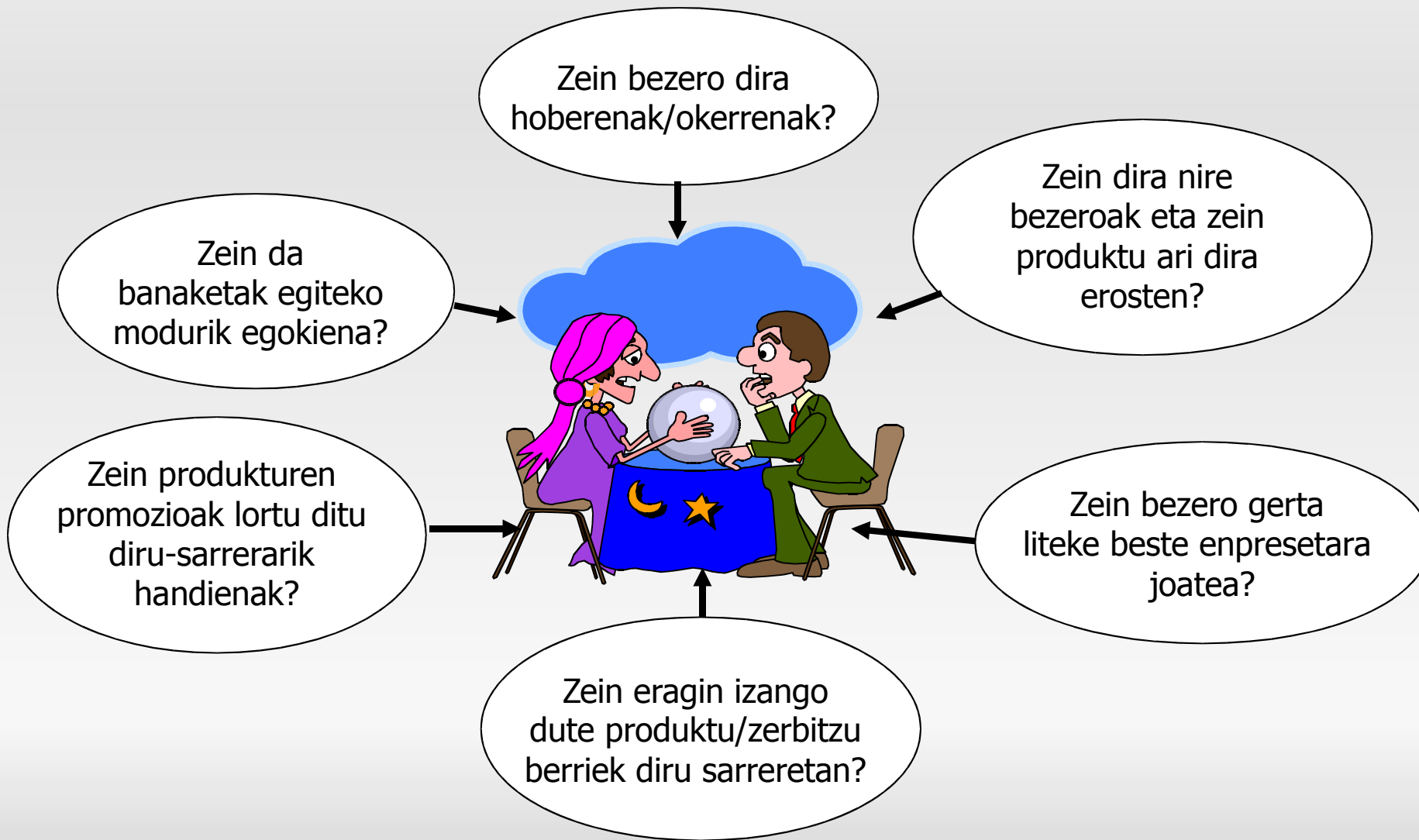
## ***(Data Warehouses)***

# Motibazioa

**OLAP.** Datu-biltegiko datu konplexuen analisia deskribatzeko erabiltzen den terminoa.

- *Gaur egun, **erabakiak hartzeko** gaitasuna duten erdi mailako zuzendariek, erabaki horiek hartzeko erraztasunak emango dizkien informazioa erabili behar dute eta honek nahikoa xehetasun behar du izan.*
- **Datuen biltegitratzeak (OLAP) (lineako prozesatze analitikoa, Online Analytical Processing)** eta datuen meatzaritzak funtzionalitate hori eskaintzen dute.

# Zergatik Data Warehousing?



# Datu-biltegiak. Definizioa

- Erabakiak hartzeko laguntza ematen duten teknologien bilduma da. Helburua: ezagutzaren arloko langileari (exekutiboa, kudeatzailea, analista...) erabaki hobeak eta bizkorragoak hartzen laguntzea.
- Datuen bilduma integratua, ez-lurrunkorra, denboraren arabera aldatzen dena eta subjektuei zuzendutakoa, kudeaketako erabakiak hartzen laguntzen duena.

W.H. Inmon (1992)

- Iturri ezberdinetatik datorren datu integratuen “multzoa” da eta dimentsio anitzeko eredu batean biltegitratua izateko prozesatua dago.

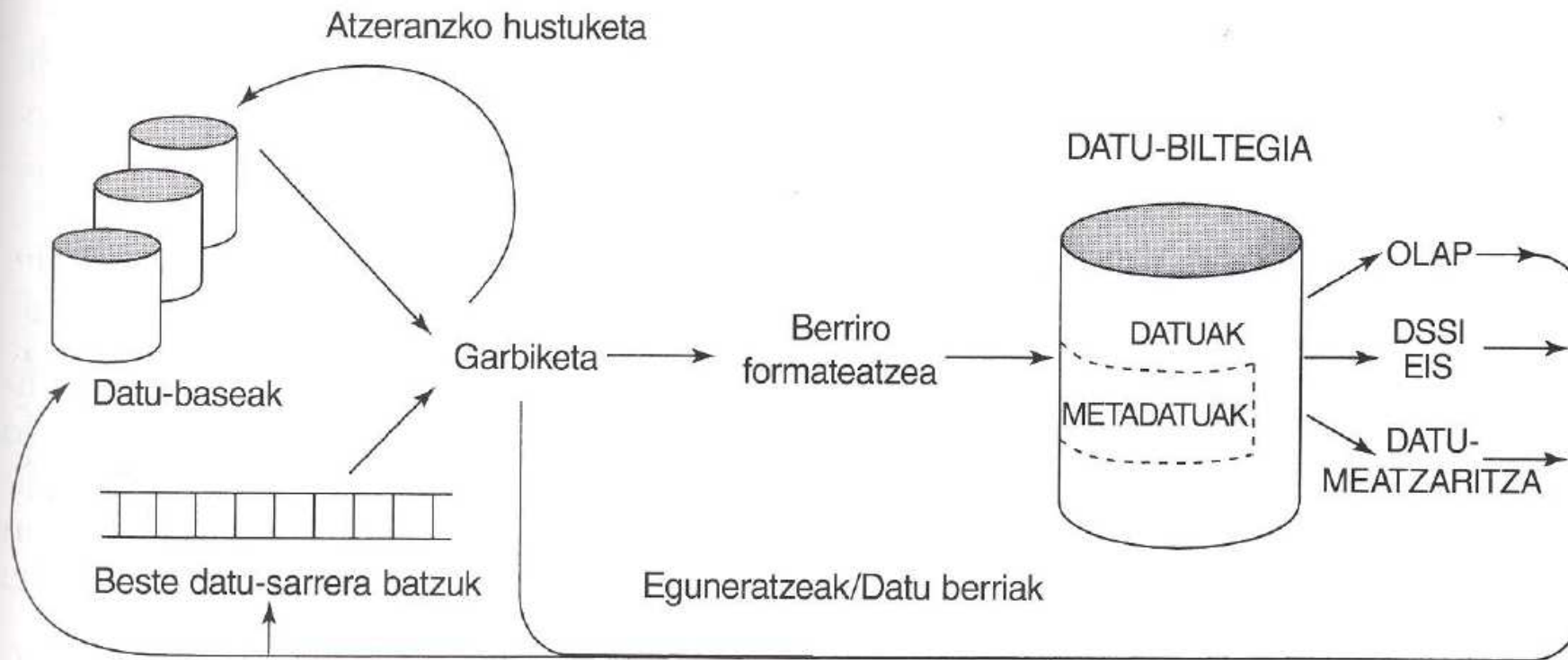
# Datu-biltegien eta DBen arteko diferentzia

- **DB.** Transakzionalak dira (informazioa berreskuratzeko kontsultez gain, txertatzeak, ezabatzeak eta eguneratzeak kudeatzen dituzte).  
**OLTP** erabil dezakete (*online transaction processing* edo lineako transakzio-prozesatzea).
- **Datu-biltegiak.** Erabakiak hartzeko aplikazioetara bideratuak. Datuen berreskuraketarako optimizatuak eta **ez** ohiko transakzioen prozesatzerako. Ez dute aldaketarik onartzen.  
**OLAP** (*online analytical processing*) erabiltzen dute.

# Datu-biltegien ezaugarriak

- Datu-eredua: Datu-biltegiak hainbat iturburuetako datu integratuetarako biltegia dira. Horiek biltegitatu aurretik, **dimentsio anitzeko datu-eredu** baten bidez prozesatu egiten dira.
- Denbora-serieak eta joera-analisiak eskaintzen dituzte.
- Ez-lurrunkorrak: informazioa askoz gutxiagotan aldatzen da; ez da denbora errealekoa. Aldizka eguneratzen da.
- Informazio-unitatea granularitate handikoa. Freskatzea noizbehinka. Normalean inkrementalki.

# Egitura kontzeptualaren ikuspegi orokorra



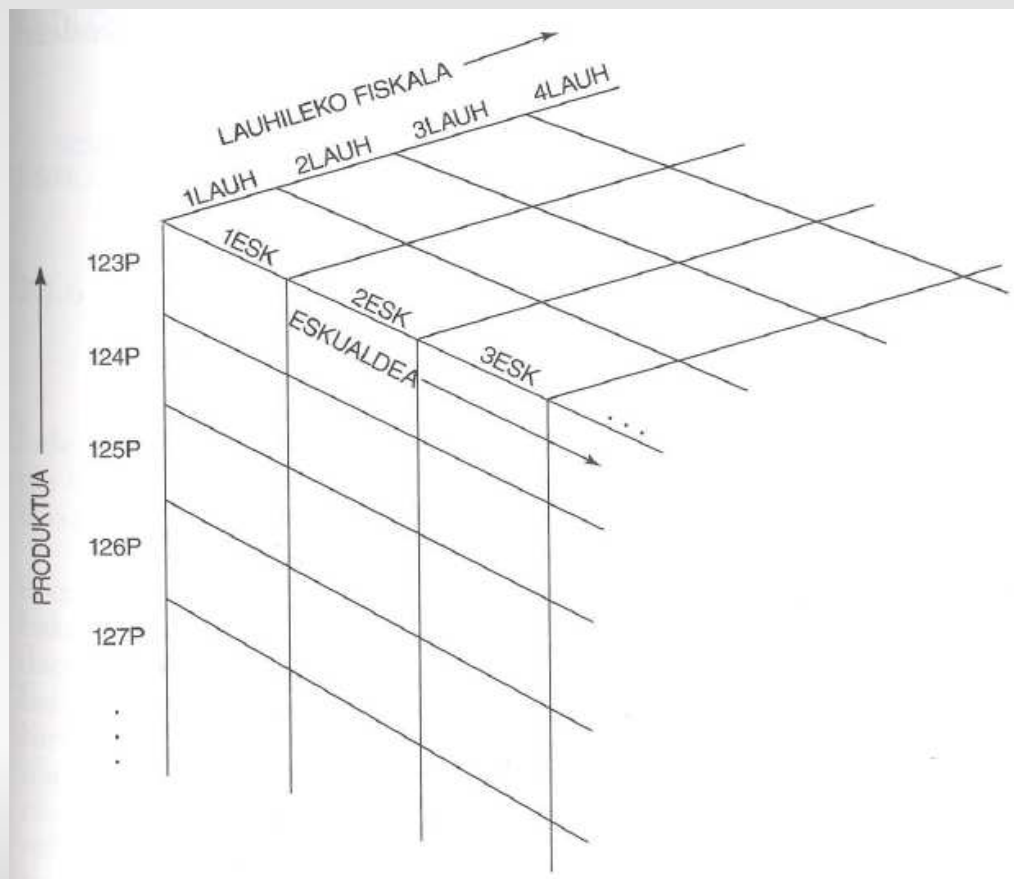
# Datu-biltegietarako datu-modelatzea

**Dimentsio anitzeko ereduak** erabiltzen dira. Hauek datuen barruko harremanak baliatzen dituzte, datu-kubo (edo hiperkubo) izeneko matrizeak datuez betetzeko.

Adibidea:

Enpresa bateko datu-biltegi baten hiru dimentsio:

- enpresaren zerga-ekitaldiak
- produktuak
- eskualdeak

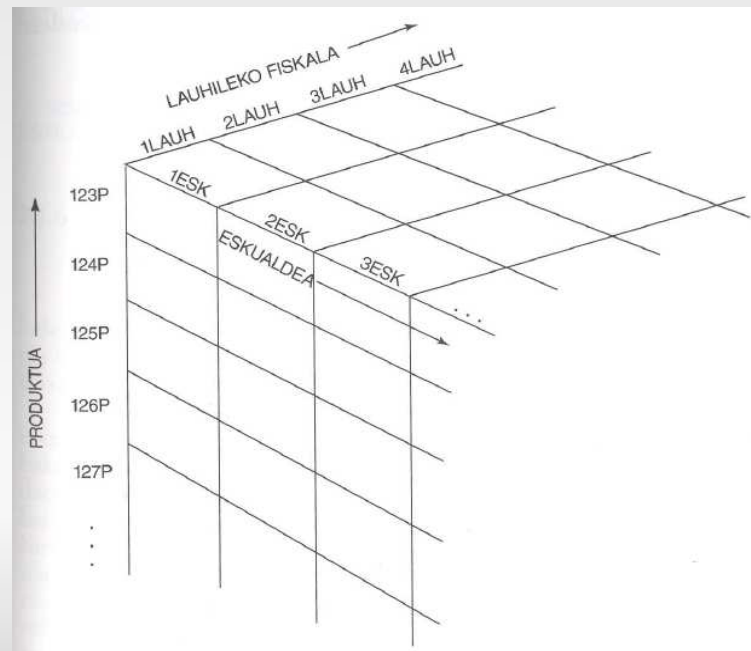




# Datu-biltegiatarako datu-modelatzea

Datuetan zuzenean egin daitezke kontsultak edozein dimentsio-konbinaziotan. Datuak erabiltzaileak hautatzen dituen dimentsioen arabera bistaratzeko tresnak badaude.

Datu-kubo batean orientazio (hierarkia) dimentsional batetik bestera aldatzea erraz egiten da **pibotatze** (edo biratze) deituriko teknika baten bidez.



# Datu-biltegietarako datu-modelatzea

Bistaratzeari buruz:

**Roll-up (trinkotzea).** Orientazioa gorantz mugitzen du, unitaterik handienak dimentsio batean multzokatuz. Adib., produktu multzoetan, edo lauhilekoak urteetan...

Biribiltze-eragiketa



**Drill-down (deskonposatu).** Granularitate xeheagoko bista eskaintzen du. Adib., eskualdeak herrietan, edo asteak egunetan...

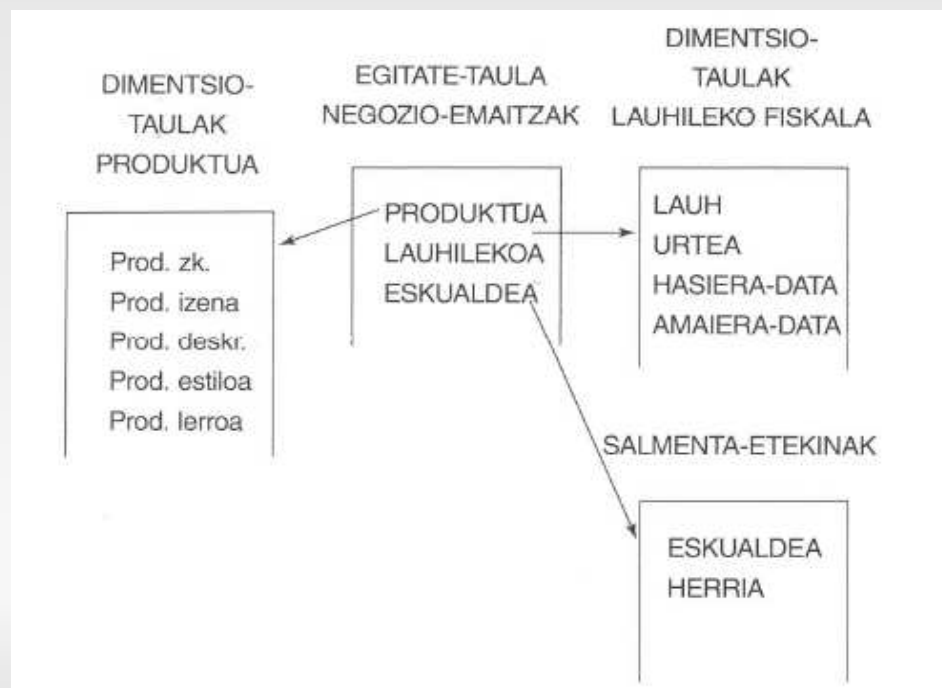
Xehadura-eragiketa



# Datu-biltegietarako datu-modelatzea

Dimentsio anitzeko ereduak bi taula mota erabiltzen ditu: dimentsio-  
taulak eta egitate-taulak.

- **Dimentsio-taula.** Dimentsioaren atributuen tuplak ditu. - **HAINBAT**
- **Egitate-taula.** Tupla multzoak. Tupla multzo bat erregistratutako egitate bakoitzerako. - **BAT**



Izar-formako eskema bat

# Datu-biltegietarako datu-modelatzea

## Dimentsio anitzeko eskemak:

- **Izar-formako eskema** (aurreko gardenkikoa). Egitate-taula bat da, dimentsio bakoitzeko taula bakarra duena.
- **Elur-malutaren formako eskema** (azpikoa). Izar-formako eskemaren aldaera zeinetan izar-formako eskemako taula dimentsionalak hierarkia batean antolatzen diren.



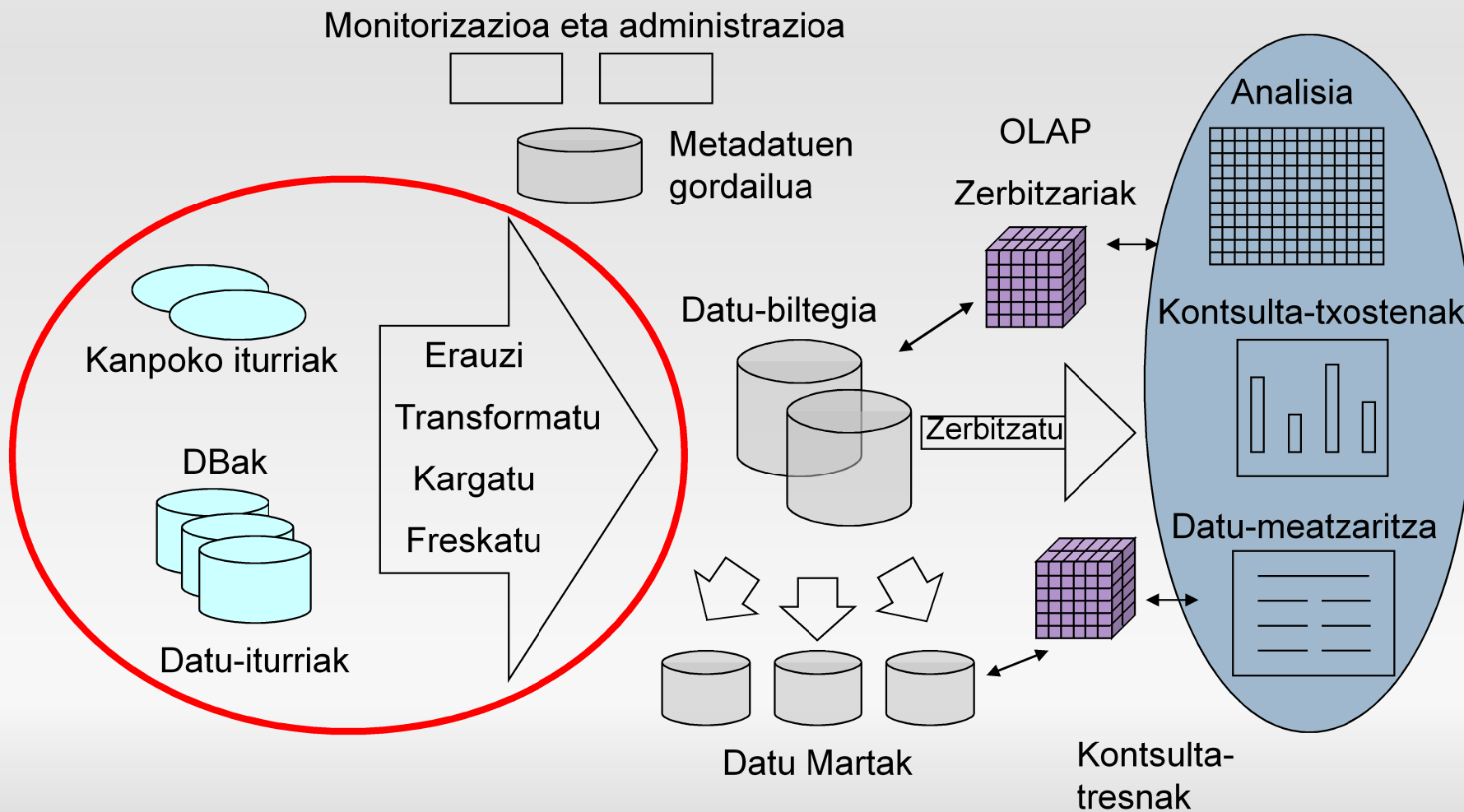
## Elur-malutaren formako eskema bat

# Datu-biltegi bat eraikitzea

Urratsak:

1. Datuak iturburu anizkoitz eta heterogeneoetatik erauzi behar dira (adib. DB, fitxategiak e.a.).
2. Datuei formatua eman behar zaie, biltegiaren barruan koherentzia izan dezaten.
  - Izenak, esanahiak eta domeinuak bateratu egin behar dira.
3. Datuak garbitu egin behar dira, baliozkotasuna ziurtatzeko.
4. Datu-iturri ezberdinetatik jasotako datuak biltegiaren datu-eredura moldatu behar dira.
5. Datuak biltegian kargatu behar dira.

# Datu-iturriak: prozesuak eta herramintak



# Datu-iturriak

## Data Sources



Operational System



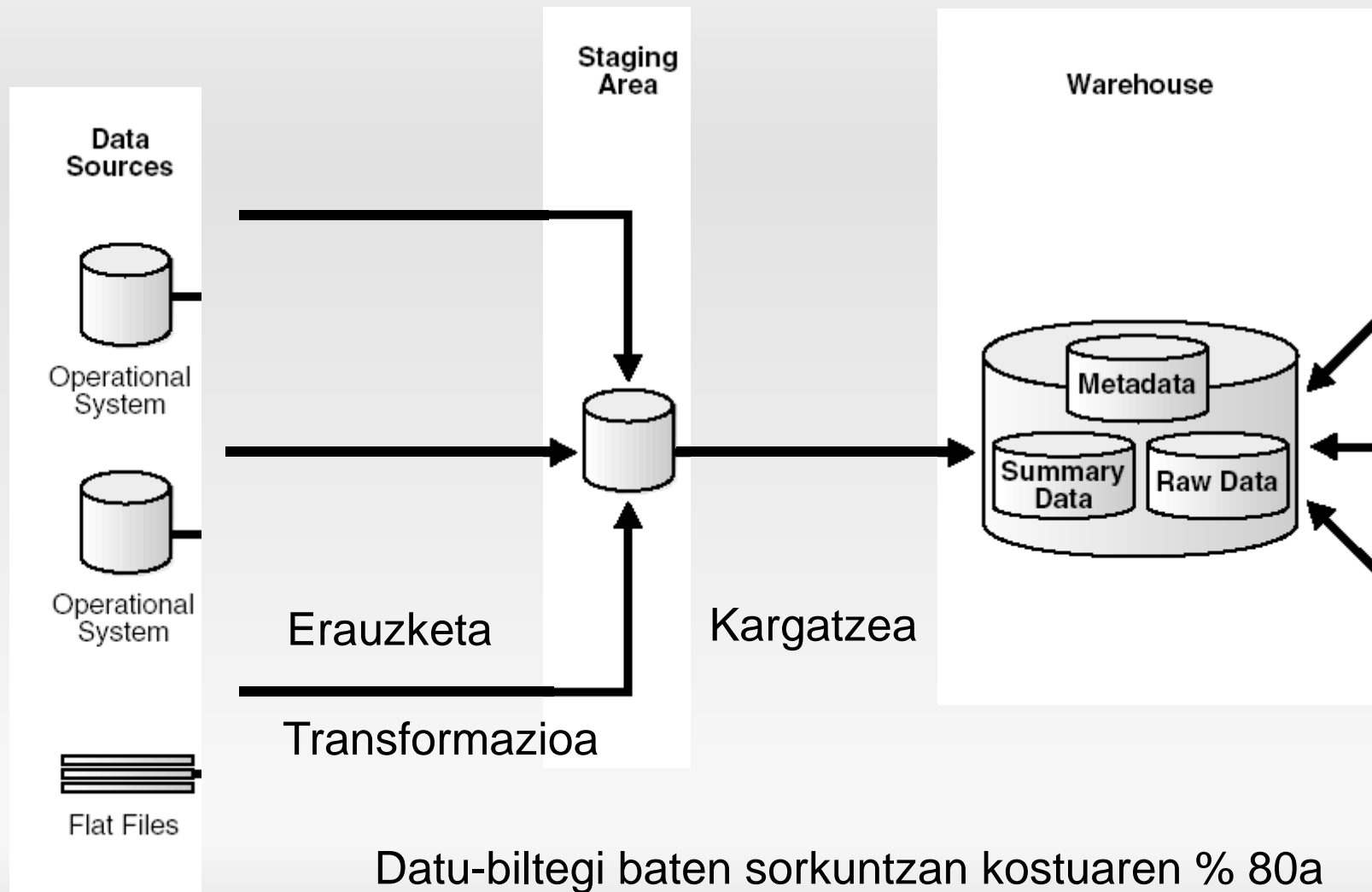
Operational System



Flat Files

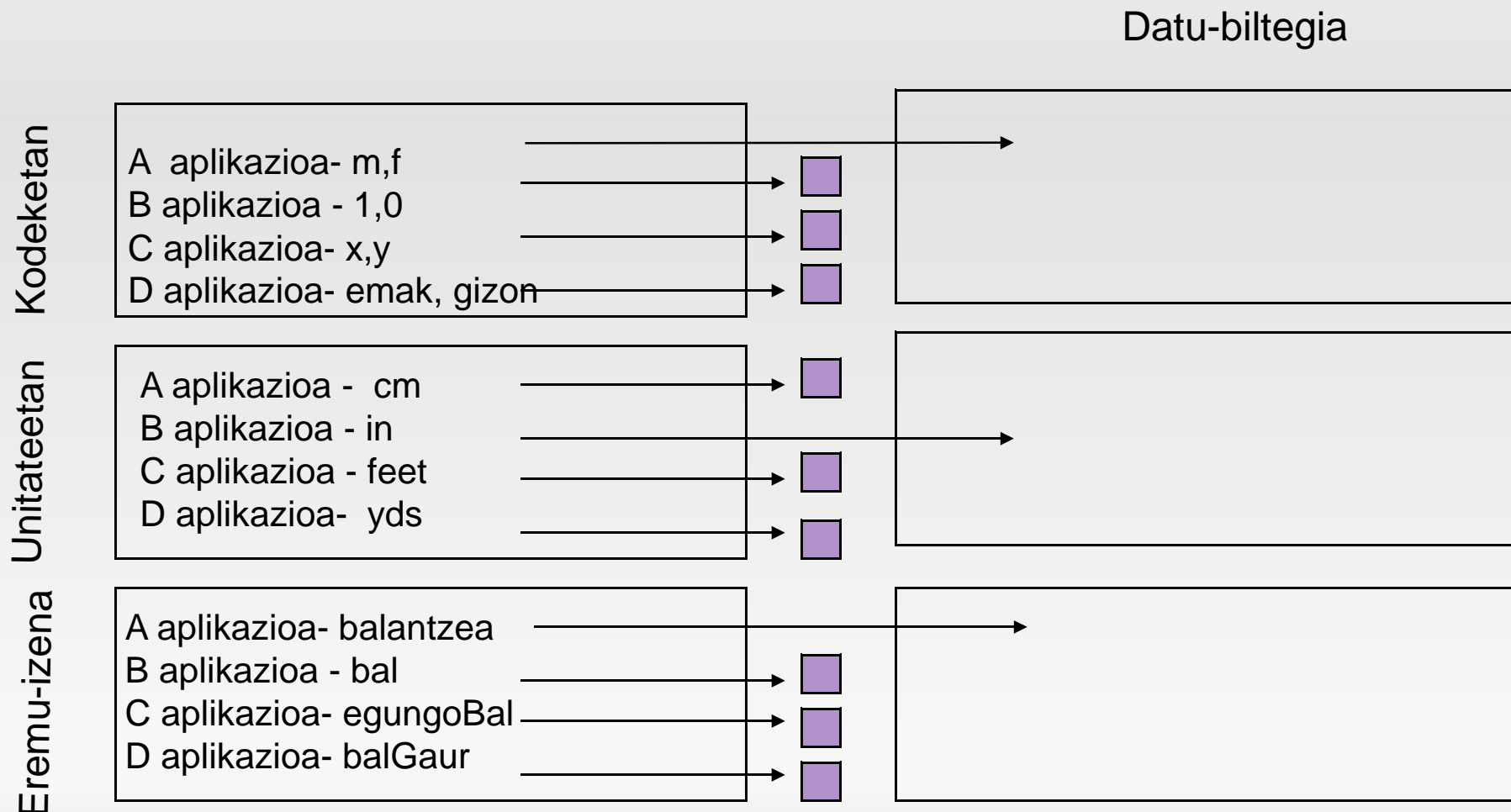
- Kanpoko datu-baseak
- Fitxategi lauak (flat)
- Datuak ohiko formatuetan:
  - Dokumentuak, fakturak, albaranak, e.a.
- Internet

# Erauzketa, transformazioa, kargatzea





# Datuen transformazioaren adibidea



# Kontsulta-zerbitzariak

Arkitektura gehienetan, datu-biltegiaren ezberdina den zerbitzari bat erabiltzen da kontsultetarako

Bi teknologia:

- ROLAP (OLAP erlazionala)
- MOLAP (dimentsio anitzeko OLAP)

# ROLAP kontsulta-zerbitzariak

- Teknologia erlazionala erabiltzen dute (Relational OLAP)
  - Teknologia erlazionala erabiliz dimentsio anitzetako datuak errepresentatu ahal izateko inplementazio-metodoak erabiltzen dituzte
  - SQL estandarraren hedapenak erabiltzen dituzte, datuen dimentsio anitzeko atzipena egiteko
- Abantaila: Estandar batean oinarritua
- Erabiliak
  - Oracle (Oracle 9i/10g)
  - IBM (DB2 y Business solutions)

# MOLAP kontsulta-zerbitzariak

- Dimentsio anitzetako teknologia erabiltzen dute (Multidimensional OLAP)
  - Datuak zuzenean matrizetan gordetzen dira
  - Kontsulta-eragiketak matrize horien gainean inplementatuta daude zuzenean
- Abantaila: ROLAP baino azkarragoak izan ohi dira
- Desabantaila: Ez dira SQL estandarrean oinarritzen