

Andmeaidad, ERP ja CRM-süsteemid IDU0010

2. loeng

Juhtimisinfosüsteemid ja andmeaidad,
multidimensionaalne modelleerimine

Eduard Ševtšenko
Eduard.sevtsenko@ttu.ee
TTÜ Informaatikainstituut

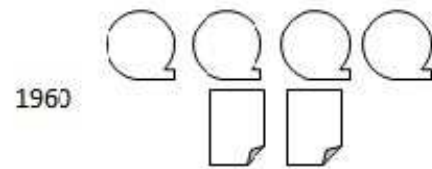
Üldist

- Õppejõud: Kontakt(Eduard.sevtsenko@ttu.ee)
 - Konsultatsioon: E 10:00 - 11:00 5A-218
K 10:00 - 11:00 ICT-645
 - kokkuleppel ka õhtused ajad
- Aine veebikodu:
<http://cloud.ld.ttu.ee/idu0010>
- Palun registreerida:
eduard.sevtsenko@ttu.ee

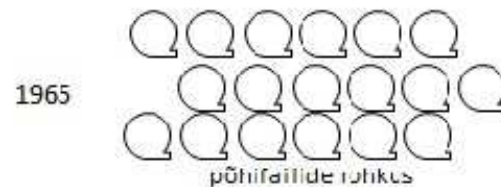
Täna sed teemad

- Juhtimisinfosüsteemide ajalugu
- Mõisted
 - andmeait (*data warehouse*)
 - andmevakk (*data mart*)
- Andmeaida modelleerimise metoodikad
 - Inmoni ülalt-alla lähenemine
 - Kimballi alt-üles lähenemine
- Andmeaida komponendid
- Sissejuhatus multidimensionaalsesse modelleerimisse
 - Faktid ja mõõdikud
 - Dimensioon ja atribuudid
 - Täheskeem (*star schema*)
- Praktikakeskkonna tutvustus ja SQL meeldetuletus

Juhtimisinfosüsteemide ajalugu



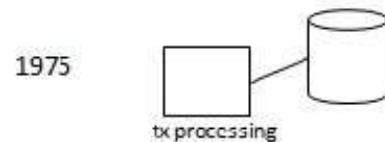
Jadapöördusfailid (*sequential access*)
Põhifailid (*master file*) ja aruanded



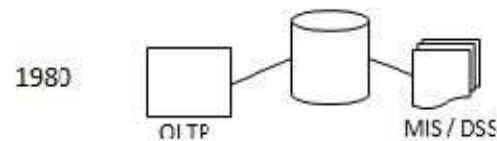
Keeruline hooldada ja arendada
andmete sünkroniseerimine
riistvaralised probleemid



Otsepöördus-ligipääs (*direct access*)
Andmebaas - ühtne andmeallikas,
mis rahuldab kõik andmetötlusvajadused



Suure jõudlusega reaalaajalised
tehingutötlussüsteemid (*Online
Transactional Processing*)
Igapäevatööga seotud operatiivandmed

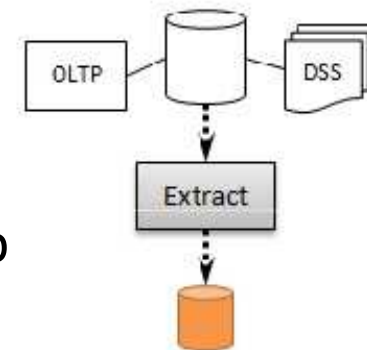


Personaalarvutid, 4. põlvkonna keeled
OLTP + juhtimisinfosüsteem
Ühtse andmebaasi paradigma

Extract e. tõmmis

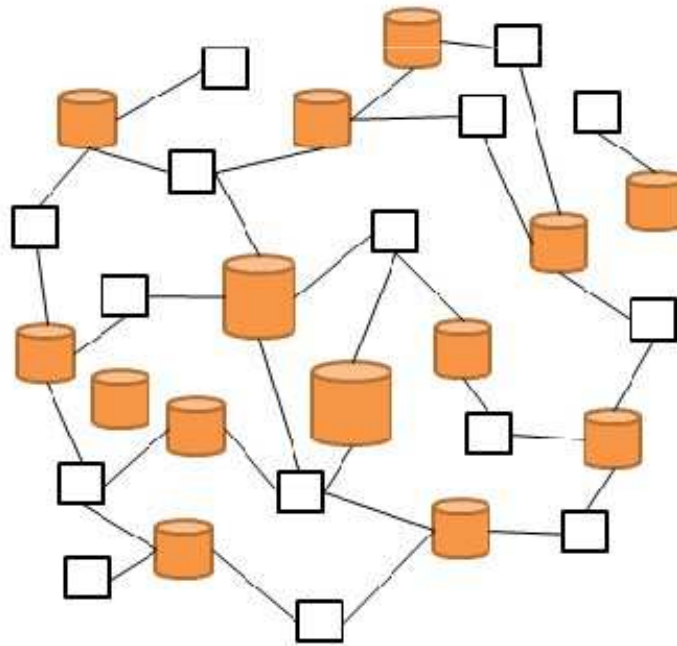
Esimene andmetöötlusprogramm

- Tekkis kohe pärast mahukate OLTP süsteemide kasutuselevõttu
- Lihtsaim andmetöötlusrakendus
 - Vaatab läbi kõik sisendandmebaasi andmed
 - Filtreerib andmed vastavalt kriteeriumitele
 - Kui leiab tingimustele vastavad andmed, kopeerib need väljundisse
- Väga laia levikuga
 - Vabastab operatiivsüsteemi andmetöötusega seotud jõudlusprobleemidest
 - Programmi kasutaja saab andmete üle kontrolli
 - vigade parandamine, tõlgendamine, grupeerimine, sorteerimine



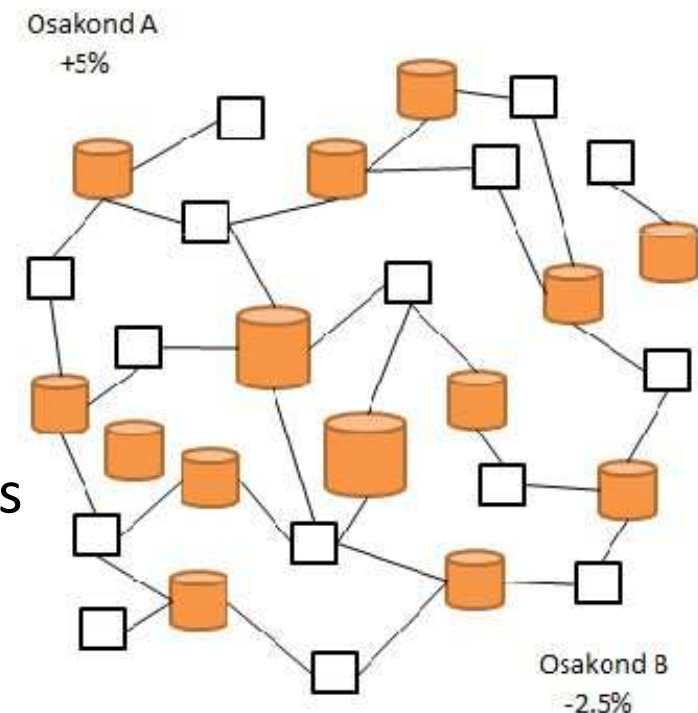
Loomulikult arenev arhitektuur

- Extract-programmide ämblikuvõrk
 - Extract, extract of extract, ...

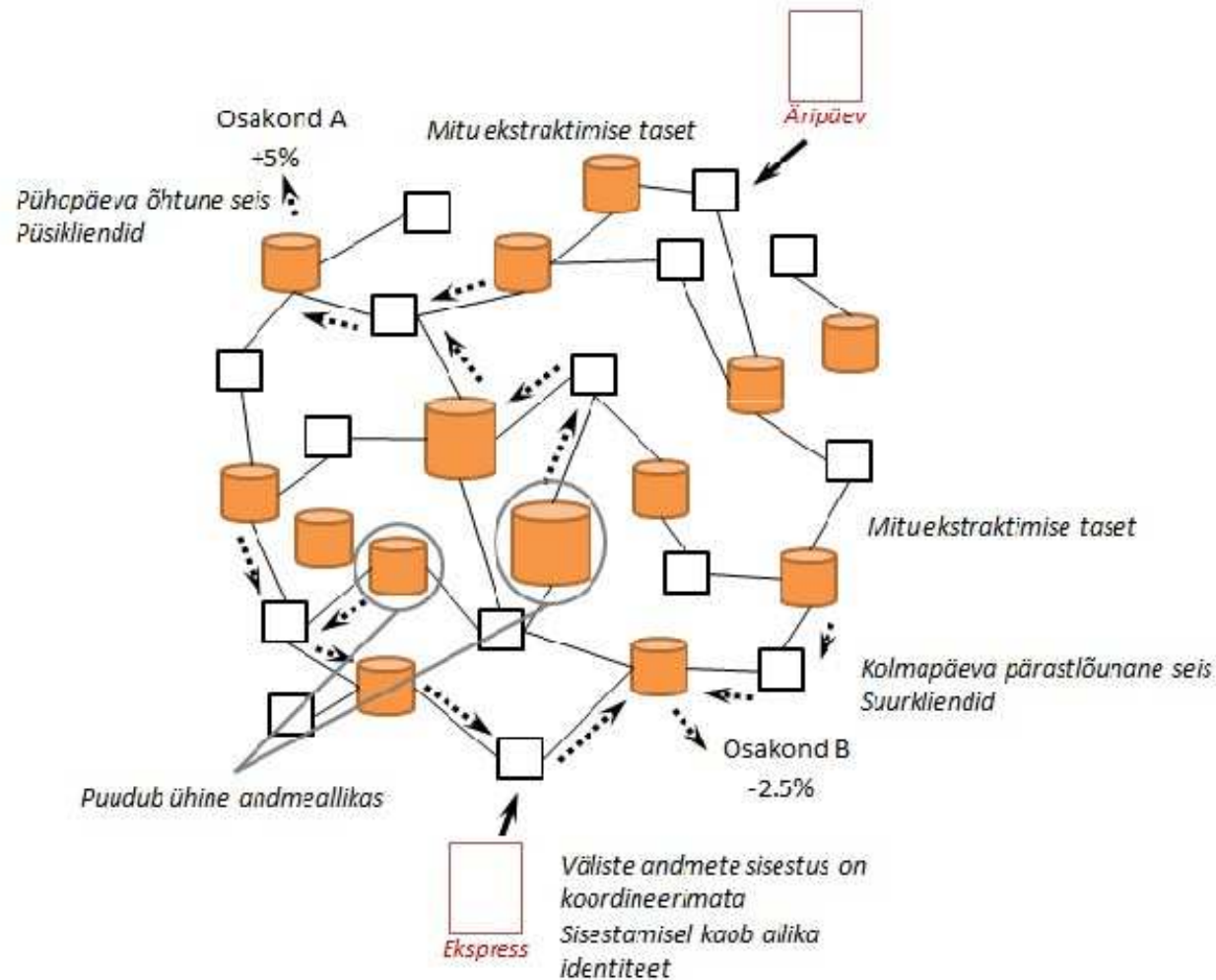


... evib mõningaid probleeme

- Andmed ei ole usaldusväärsed
 - Kuidas arvutada ettevõtte koondtulemust?
 - Juhtkond teeb otsuse isiksuste ja poliitika põhjal
- Usalduskriisi põhjused
 - Puudub ühtne ajaline "stardipakk"
 - Erinevus andmetöötamise algoritmides
 - Ekstraktimise tasemed
 - Väliste andmete haldamine
 - Puudub ühine andmeallikas

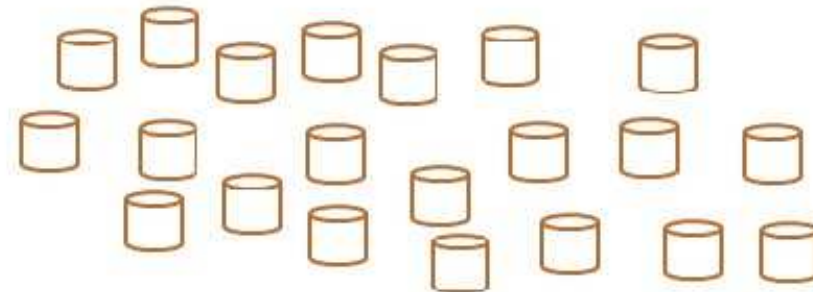


Usalduskriisi põhjused

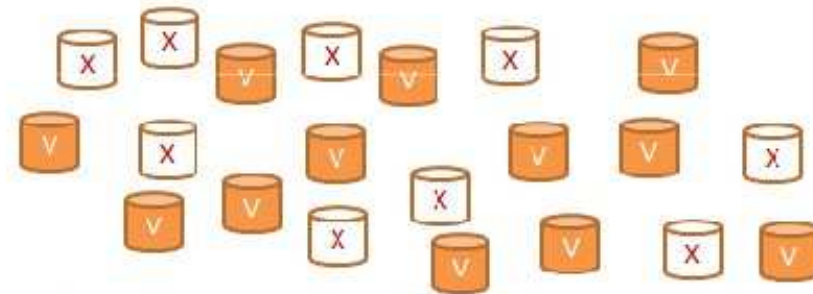


Produktiivsus?

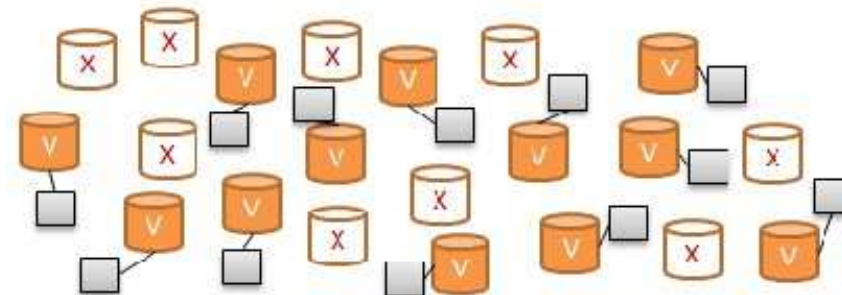
- Andmeallikate tehnoloogiline mitmekesisus
 - MySQL, PostgreSQL, Exceli tabelid
- Andmete sisuline mitmetähenduslikkus ja ebakooskõla
 - Väljad "Balance", "Amount"
- Andmetele ligipääs on kulukas ja aeganõudev
- Protsess ei ole korduvkasutatav: iga laiaulatusliku aruande koostamisel samad probleemid



Koosta aruanne, mis hõlmab tervet ettevõtte ajalugu



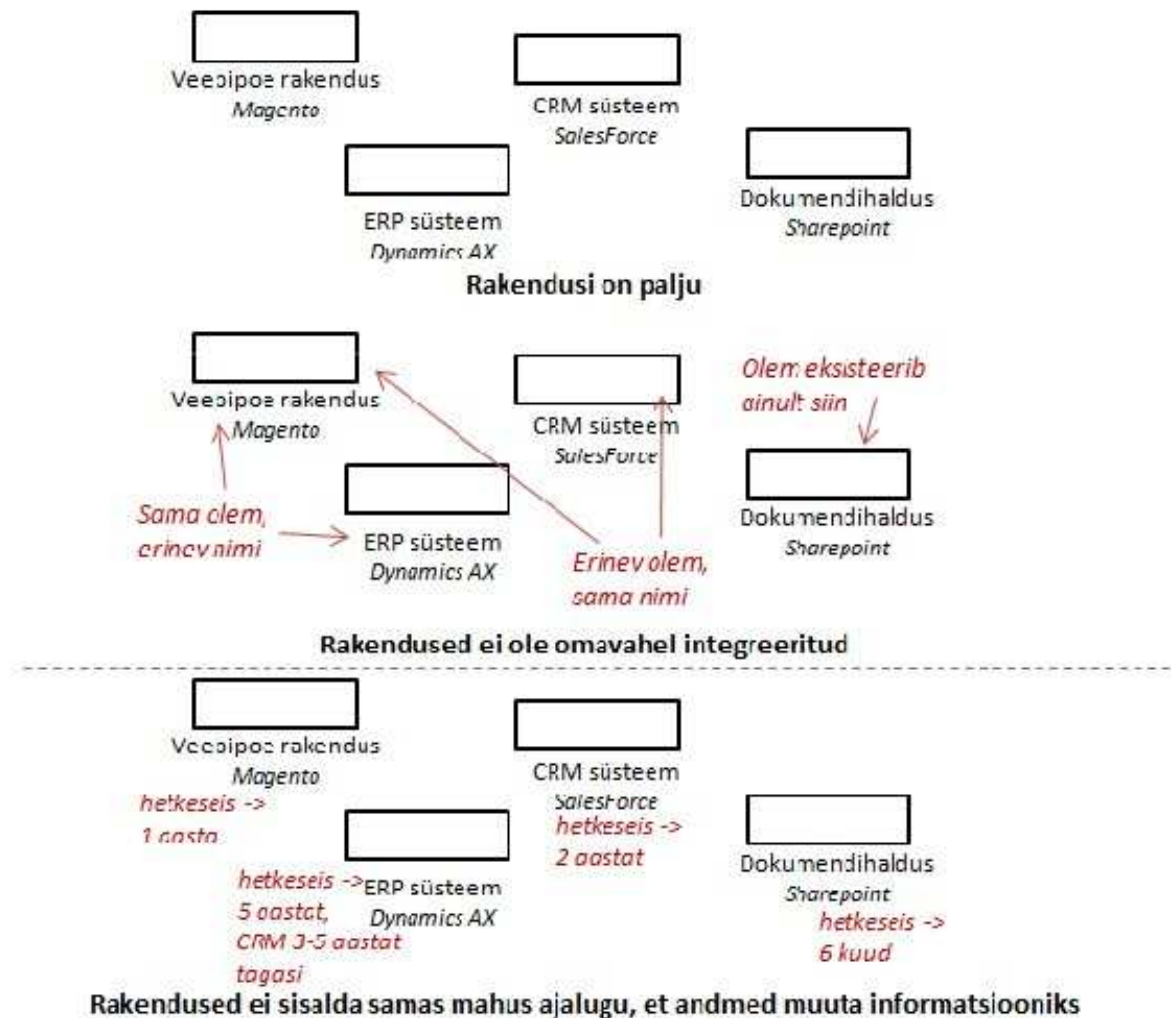
Andmete leidmiseks tuleb läbi vaadata väga palju faile/andmebaase



Palju kohandatud Extract-programme, tehnoloogiline keerukus

Andmed informatsioon

- Kuidas on klientide ostueelistused muutunud viimase 5 aasta jooksul?



Kas tuleb tuttav ette?

Ühe mehe vead on teise mehe andmed

- Faktidele on raske tugineda, kui:
 - Meil on firmas tonnide viisi andmeid, aga me ei pääse neile ligi
 - Mulle ei sobi valmis aruanded, ma soovin andmeid mitut-setut moodi filtreerida ja kombineerida
 - Kuidas on võimalik, et kaks osakonda suudavad koosolekul sama finantsnäitaja kohta esitada erineva numbri?
 - IT, anna meile ligipääs andmebaasile, küll me ise vaatame, kuidas sealt andmeid kätte saame
 - Näita mulle ainult seda, mis on oluline!

Muudame lähenemist

Andmete "kaks maailma"

Primitiivsed e. operatiivandmed

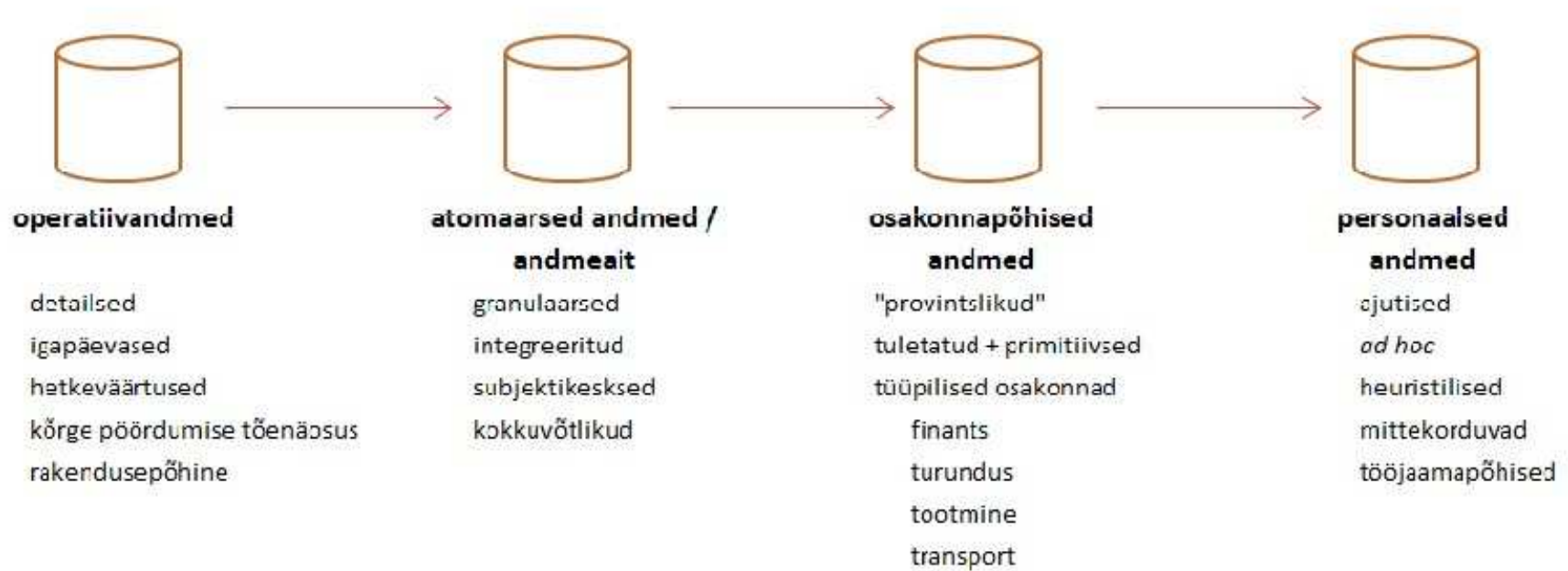
- rakendusespetsiifilised
- detailed
- täpsed hetkeväärtused
- kasutatakse regulaarselt
- infovajadused on eelnevalt defineeritud, kaardistatavad ning ajas vähemuutuvad
- tundlikud süsteemi jõudluse suhtes
- tehingupõhine
- nõuetepõhine süsteemiarenduse elutsükkel
- kõrge pöördumise tõenäosus
- kasutusmuustrilt pöördumine üksiku andmemelemendi või kirje poole
- valdav toiming on sisestamine ja muutmine (CU)
- on muudetav, uuendamisel probleemiks andmete omandisuhe
- mitteliase sisuga
- staatiline struktuur, muutuv sisu
- toetab igapäevategevusi, et organisatsioon üldse töötaks ja protsessid toimiks

Tuletatud e. otsustusandmed

- olemi e. andmesubjekti kesksed
- kokkuvõtlikud, vajadusel täpsustatud
- esitavad väärtuste muutusi ajaloo vältel, tõmmised
- kasutamine heuristiline (ei ole prognoositav)
- infovajadusi ja nõudeid ei saa *a priori* defineerida, muutuvad kiiresti ja prognoosimatult
- ei ole jõudluskriitiline
- analüüsipõhine
- primitiivsetest erinev elutsükkel - andmetepõhine
- madal pöördumise tõenäosus
- kasutusmuustrilt pöördumine andmehulkade (sadade tuhandete kirjete) poole
- valdav toiming on andmete lugemine (R)
- on arvutatav, vigade parandamine ei tekita probleeme andmete omandis/vastutuses
- struktuurilt paindlik ja liiane
- toetab juhtimistegevusi: möödunu analüüs ja tuleviku prognoosimine

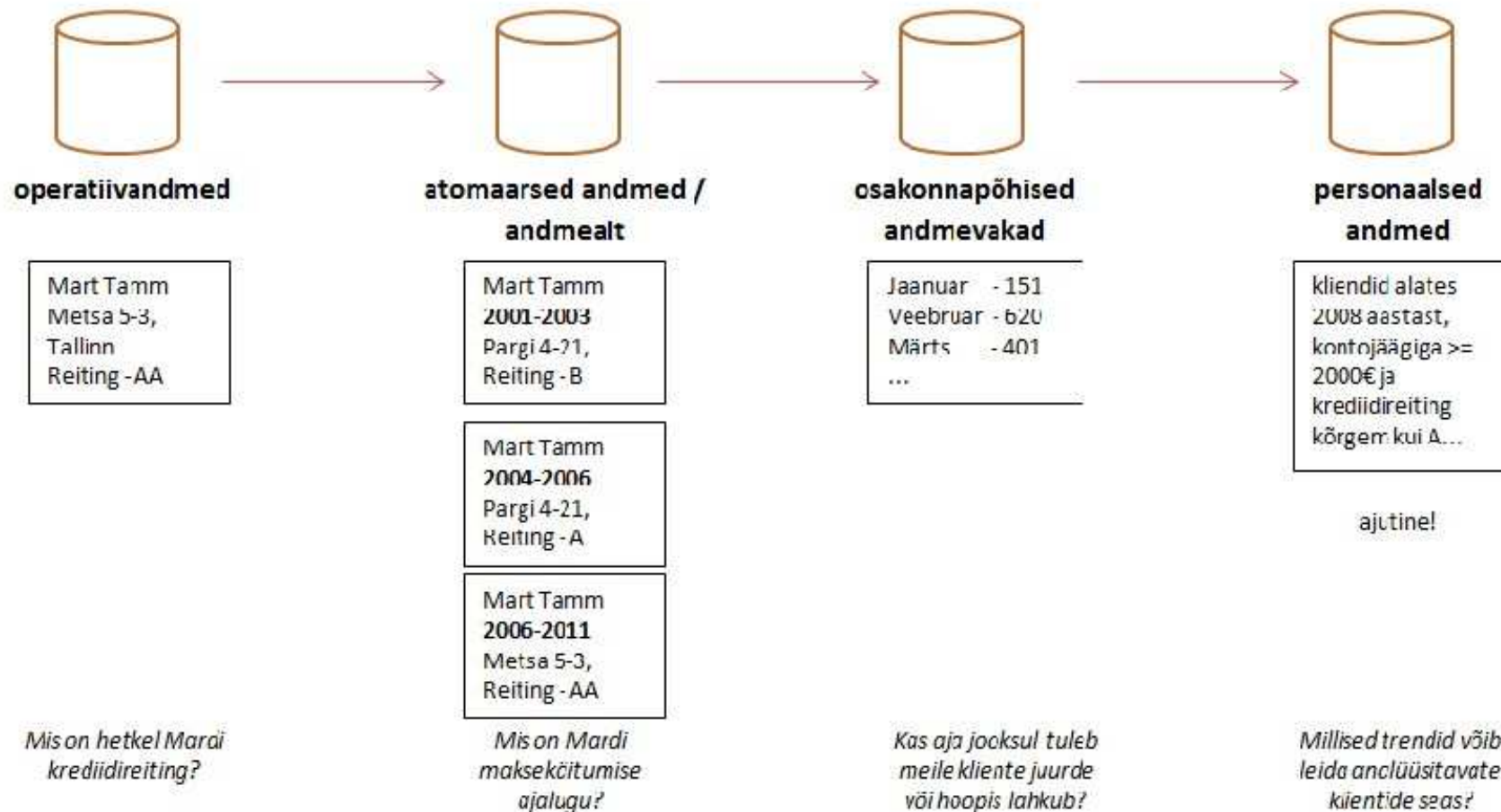
Andmetöötuse tasemed

The Architected Environment

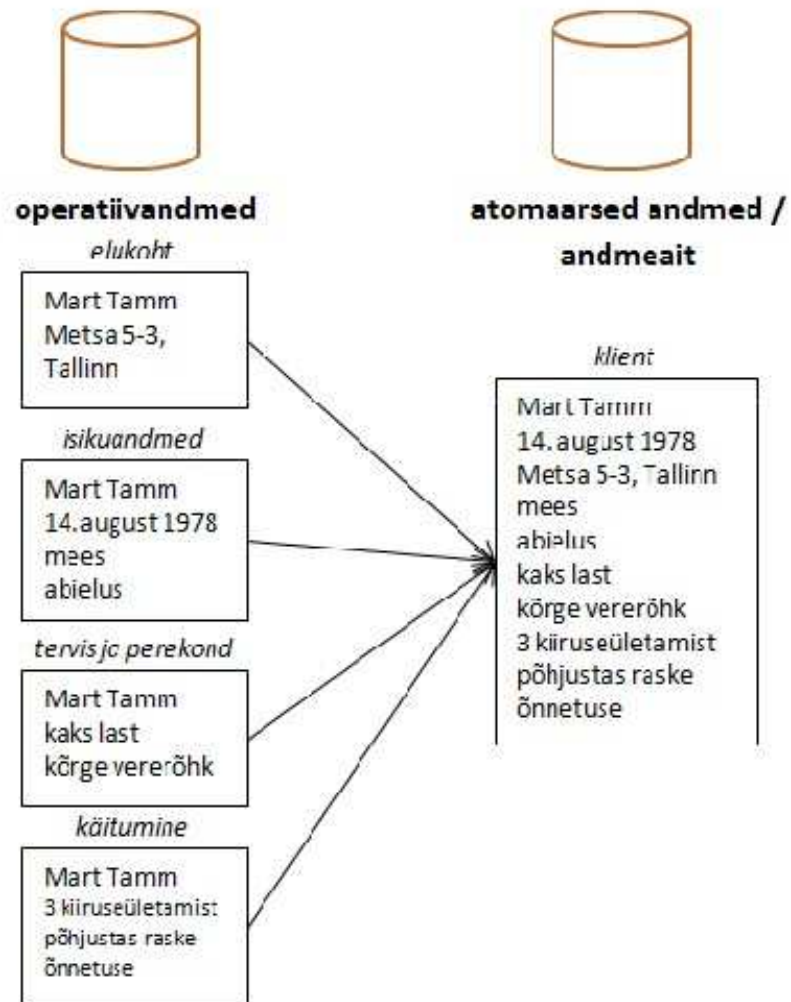


Päringud ja küsimused sõltuvad tasemest

Lihtne näide - klient



Andmeait integreerib andmed



Andmeait on subjektipõhine

Operatiivandmed

- Sõiduk
- Reis
- Tervis
- Hoone

on rakendusepõhised

Otsustusandmed

-  Klient
-  Poliis
-  Kahjunõue
-  Hüvitis

koondavad subjekti kohta käiva

Andmeaida eesmärgid

- Andmeait peab muutma organisatsiooni andmed lihtsalt kasutatavaks, st. hõlbustama andmete teisendamst informatsiooniks:
 - Andmed peavad olema lõppkasutajale mõistetavad (meta-andmed ja seosed).
 - Andmeid peab saama erinevatel viisidel filtreerida ja kombineerida (nn. slicing and dicing).
 - Vastus päringutele peab saabuma kiiresti.
 - Tüüpprobleem: "Meil on palju andmebaase, aga me ei saa sealt (ilma IT-spetsialistita) andmeid kätte."
- Andmeidas paiknevad andmed peavad olema kooskõlalised (consistent) ja usaldusväärsed.
 - Andmed peavad olema puhastatud ja korrektselt seostatud.
 - Andmed erinevate äriprotsesside kohta peavad olema kooskõlas (nt. tellimused ladu arved maksed).
 - Meta-andmed peavad olema ühetähenduslikud ja intuiitiivsed: mõiste "käive" omab igal pool sama sisu.
 - Tüüpprobleem: koosolekul vaidlus müügijuhtide vahel: ühe arvutuste järgi oli viimase kuu käive 10M, teise arvates 11,5M.
- Andmeait peab kohanduma ärivajaduste järgi
 - Muutuste haldus tuleb juba sisse projekteerida - klassikaline kose-mudel andmeaitade arenduse puhul ei toimi, kuna nõudmised on pidevas muutumises.
 - Paradoks: ükski muudatus ei tohi olemasolevaid andmeid muuta selliselt, et tekkiks "ajaloo ümberkirjutamine".
- Ligipääs andmeidas hoitavatele andmetele peab olema reguleeritud
 - Andmeaita on kogutud informatsioon organisatsiooni erinevate valdkondade kohta - seda tuleb kaitsta.
 - Mis on ärisaladus: kliendid, klientide andmed, kaubad, kaupade hinnad, allahindlused ja kampaaniad, marginaalid?
- Andmeait peab toetama otsustusi (decision support)
 - Otsustuste tegemiseks on vaja korrektset ja usaldusväärset informatsiooni möödunu kohta.
 - Andmelao kasutamine ei ole sageli möödapääsmatu (vs. operatiivsüsteem), seega tuleb see ärikasutajatele edukalt 17 "maha müüa".

Andmeaida koostamise metoodikad

Inmoni ülalt-alla lähenemine

- Andmeait on üle-organisatsiooniline andmebaas - "ühe tõe allikas"
- Andmeaida koostamisel lisatakse sinna subjekte (nt. klient, kaup, müük) ning nende kohta käivaid detailseid tehinguid (arve, tellimus vm)
- Andmeait on andmestruktuurilt normaliseeritud (3NF)
- Andmevakad saavad oma lähteandmed andmeaidast, denormaliseerimine ja summeerimine toimub andmevaka tasemel
- Andmeait on:
 - teemapõhine (*subject-oriented*)
 - andmed on koondatud ümber teemade (nt. müük, tootmine), mitte operatiivsete tegevuste (nt. tellimuste vastuvõtt)
 - püsiv (*non-volatile*)
 - andmeaita lisatud andmed on lõplikud ning ei muutu
 - integreeritud (*integrated*)
 - andmed on ühilduvad, samas vormingus ja ühetähenduslikud
 - ajateisendlik (*time-variant*)
 - andmeaida põhjal on võimalik liikuda "ajas tagasi", st. lisaks tehingutele talletatakse info ka püsiandmete (nt. klient, kaup) muutuste kohta.
 - kui andmeait vastab ajateisendlikkuse tingimusele, on andmeaida põhjal võimalik taastada operatiivandmebaasis sisalduv informatsioon

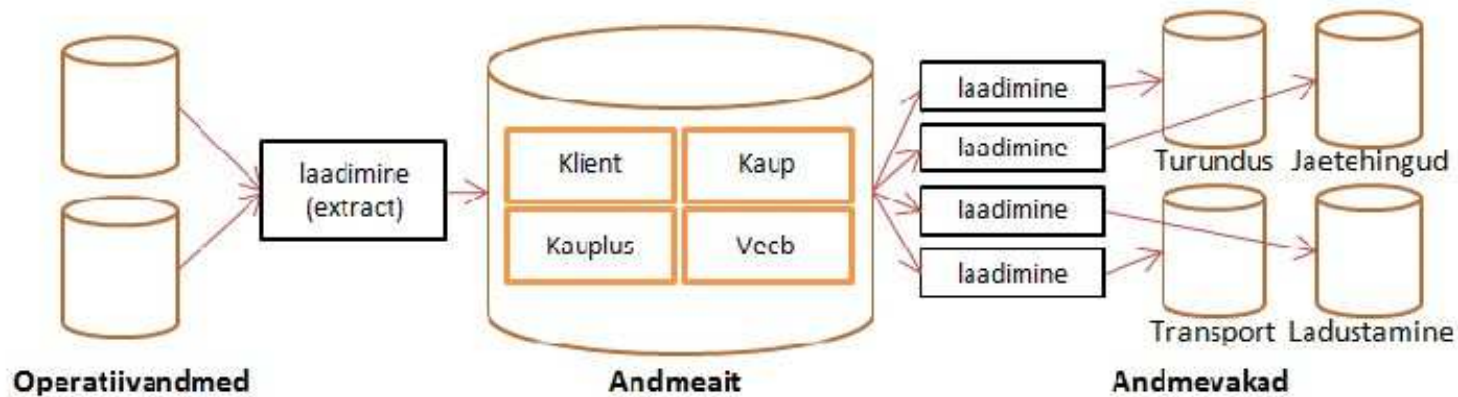
Andmeaida koostamise meetodikad

Kimballi alt-üles lähenemine

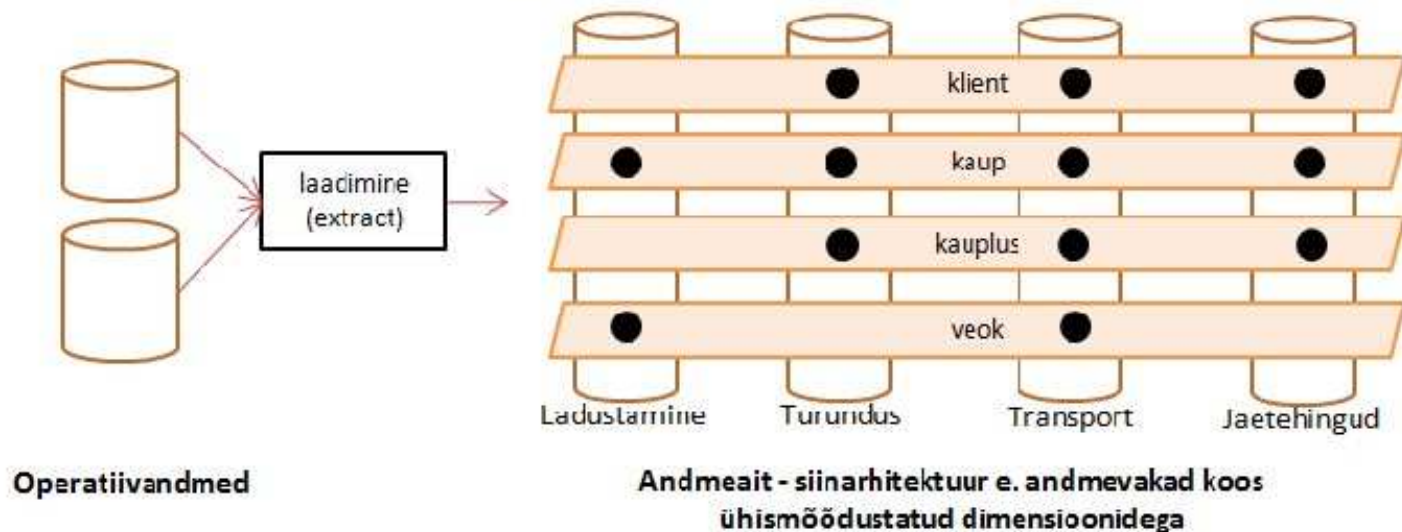
- Andmeaida koostamisel alustatakse üksikute valdkonnapõhiste andmevakkade loomisest, mis rahuldavad erinevate osakondade infovajadusi
- Andmevakkad koosnevad dimensioonidest ja faktidest
 - Dimensioon tähistab kõiki subjekti (nt. klient, kaup) kohta käivaid andmeid
 - Faktid võivad olla sõltuvalt vajadusest detailsed või ka summeeritud
- Andmevaka loomisel ühtlustatakse subjektid dimensioonideks, mida saab järgmistes andmevakkades kasutada - nn. andmeaida siin-arhitektuur (*data warehouse bus architecture*)
- Siinarhitektuur võimaldab *drill across* teisendust andmevakkade vahel, mille ühendavaks lüliks on ühtlustatud dimensioonid
 - Nt kui jaemüük ja turundus on ühendatud dimensioonidega Klient, Kauplus ja Kaup, saab vaadata, kuidas turundustegevus mõnes poes mõjutab toodete jaemüüki jm.

Metoodika vs. andmeida arhitektuur

Inmoni ülalt-alla metoodika



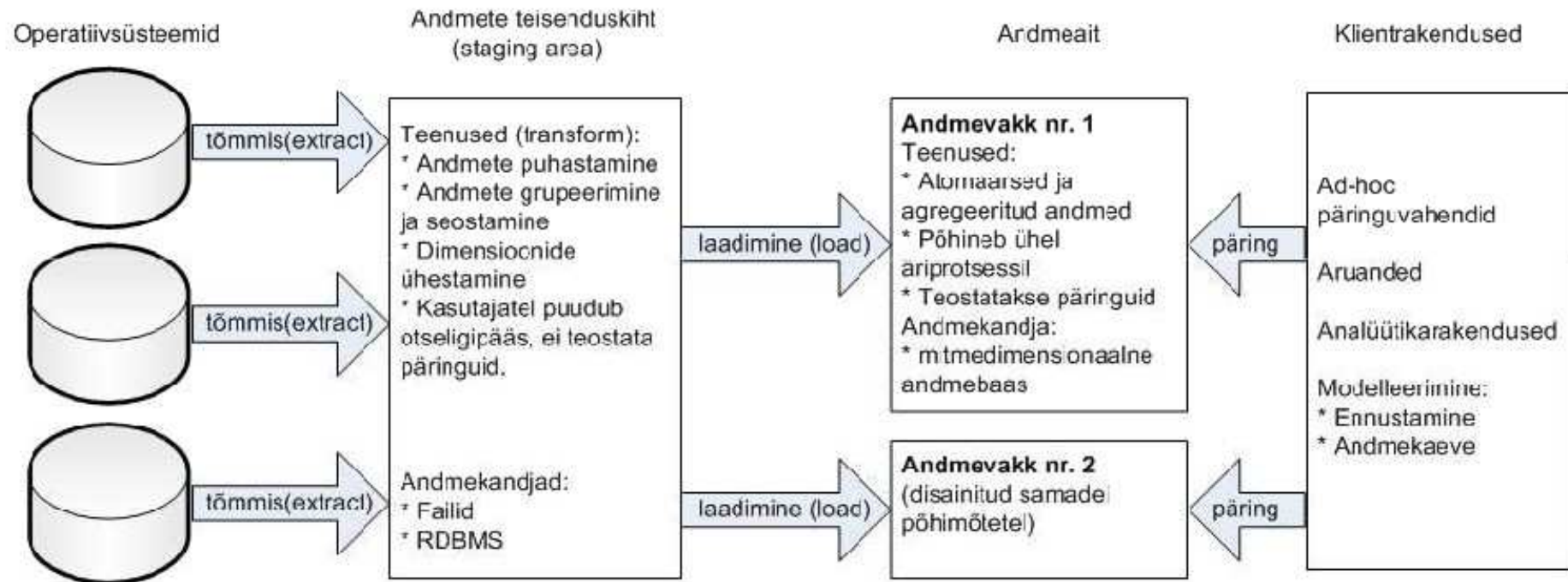
Kimballi alt-üles metoodika



Jätkame Kimballi metoodika põhiselt

Andmeaida komponendid

Kimballi metoodika põhjal



Andmeida komponendid

Kimballi metoodika põhjal

- Operatiivsüsteemide andmebaasid (OLTP - OnLine Transactional Processing database)
 - Tavaliselt andmebaas, milles hoiab andmeid ERP, veebiportaal vm. rakendus.
 - Reeglina asuvad OLTP-baasid andmelaost väljaspool, võimalikud on heterogeensed kooslused (ERP kasutab Oracle't, veebiportaal MySQL'i, ligipääsu- ja turvasüsteem MSSQLi jne)
- Andmete teisenduskiht (*staging area*)
 - Andmete ekstraktimine puhastamine, töötlemine (grupeerimine ja denormaliseerimine) ja seostamine - ETL = Extract, Transform & Load
 - Kui andmed pärinevad heterogeensetest allikatest, siis luuakse kõiki andmeallikaid katvad ühistabelid (nt. klienditabel ERPst ja veebiportaali baasist)
- Andmeait (*data warehouse*)
 - Salvestab eelnevate töötluste tulemuse: kontrollitud, kooskõlalised ja seostatud andmed
 - Iga äriprotsessi kohta on koostatud eraldi andmevakk (*data mart*)
 - Tüüpiliselt hoitakse andmeid mitmedimensionaalsetes andmebaasides (OLAP = OnLine Analytical Processing database)
 - Andmevaka vastu käivitavad päringuid lõppkasutajad
- Klientrakendused on tööriistad, millega on mugav andmeaitadest andmeid vaadata.
 - Ad-hoc päringuvahendite näiteks on MS Excel, ProClarity vm
 - Klientrakenduste kihti kuuluvad ka (staatilised) aruanded ja töölaud, mis sobivad vahetult lõppkasutajale ning kuvavad olulist informatsiooni ilma ajamahuka parameetrite seadistamiseta

MultiDimensionaalne andmeMudel

Dimensional data model

- MDM koosneb faktidest ja dimensioonidest
- Fakt (*fact*) hoiab endas äriprotsessi toimumist kirjeldavaid arvulisi mõõdikuid (*measure*)
 - Mõõtmist teostatakse alati mingis ärilises kontekstis
 - igapäevaselt müüdüd kaupade kogused ja summad
 - Mõõtmisel fikseeritakse teda kirjeldavad tunnused e. dimensioonid
 - Klient, Kaup ja Kuupäev
 - Dimensioonid määravad mõõtmise konteksti (mille suhtes?)
 - Mõõtmistulemust kirjeldavad arvulised väärtused
 - Müüdüd kogus ja Summa.

Müügifakt
Kuupäevakood (FK)
Kaubakood (FK)
Kliendi kood (FK)
Müüdüd kogus
Summa

Dimensioon

Dimension

- Dimensioon on mõõtmistulemust selgitav tekstiline kirjeldus, mis tähistab tüüpiliselt subjekti
- Dimensioon sisaldab mõõdetava tunnuse kohta kõiki olulisi atribuute
 - nt. kauba puhul: nimetus, mudel, värv, pakend
- Igal dimensioonil on primaarvõti, mis seob selle faktitabeliga
- Atribuutide nimed peavad olema üheselt mõistetavad andmeida lõppkasutajale
- Andmevakale annavad väärtuse dimensiooni atribuutide rohkus ja sisukus.
 - Küsimusele: "Kui palju müüdi eelmisel kuul kilepakendis piimatooteid?" saab vastata ainult siis, kui dimensioonitabel sisaldab vastavaid atribuute

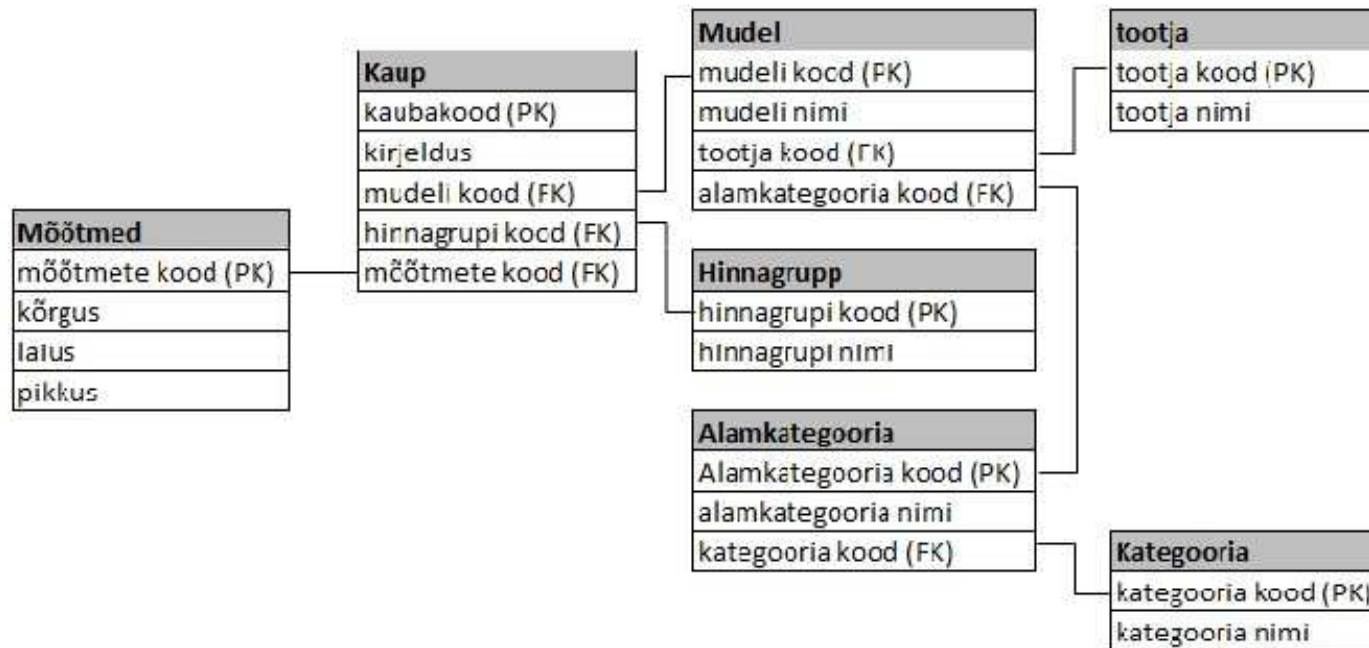
Kauba dimensioon
Kauba kood (PK)
Kauba nimetus
Mudel
Tootekategooria
Pakendi suurus
Pakendi tüüp
Eluiga
Mass
...

struktuur

näidis-
andmed

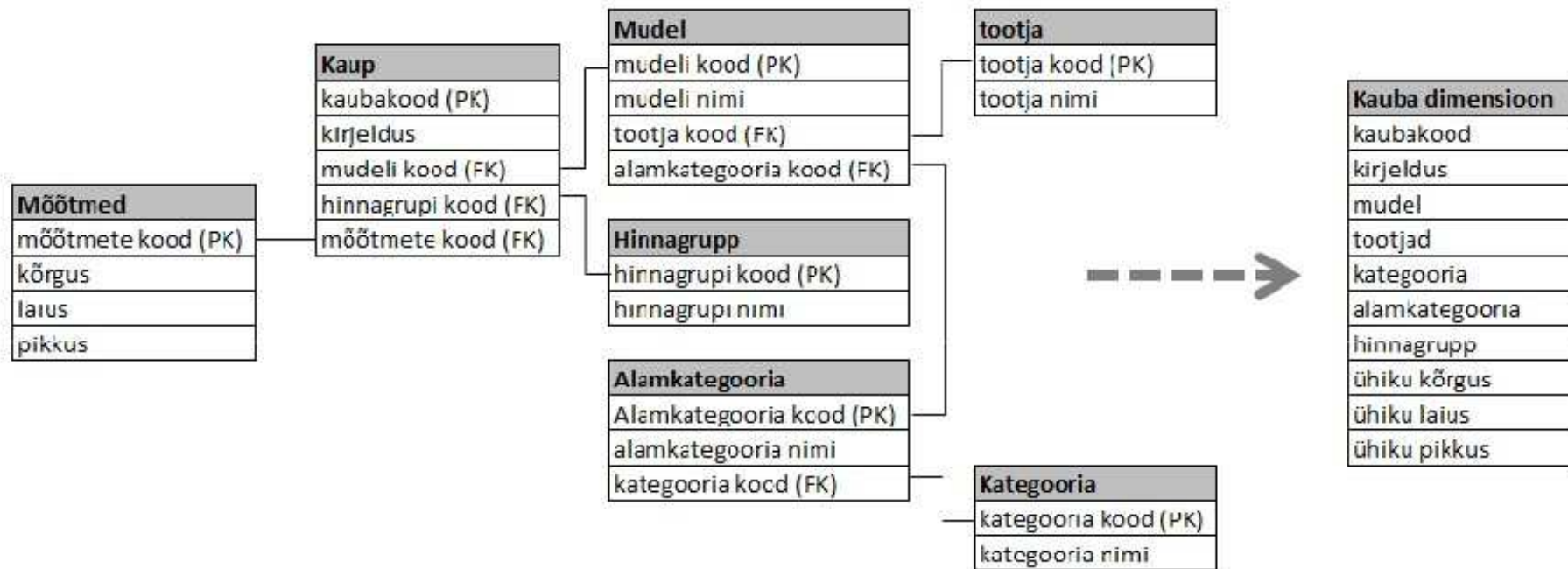
Kauba kood	Kauba nimetus	Mudel	Tootekategooria	...
E24664	Tahvelarvuti Samsung	Galaxy tab	Pihuseadmed	...
F43674	Tahvelarvuti iPad	ipad 2	Pihuseadmed	...
r94642	telef LG	42LH5U	Telefid ja displeid	...
Q64383	MP3 mängija	ipod	MP mängijad	...
X32464	MP3 mängija IPoc	Mini 6G	MP mängijad	...

Kuidas tekivad dimensioonid?



Denormaliseerime andmed

ETL programmi abil



```

SELECT K.kaubakood, K.kirjeldus
      , MD.mudeli_nimi AS mudel, T.tootja_nimi AS tootja
      , TK.kategooria_nimi AS kategooria, TAK.alamkategooria_nimi AS alamkategooria
      , HG.hinnagrupi_nimi AS hinnagrupp, ...
FROM kaup K
LEFT JOIN mudel MD ON MD.mudeli_kood = K.mudeli_kood
LEFT JOIN tootja T ON T.tootja_kood = MD.tootja_kood
LEFT JOIN alamkategooria TAK ON TAK.alamkat_kood = MD.alamkat_kood
LEFT JOIN kategooria TK ON TK.kategooria_kood = TAK.kategooria_kood
LEFT JOIN hinnagrupp HG ON HG.hinnagrupi_kood = K.hinnagrupi_kood
LEFT JOIN mõõtmed MM ON MM.mõõtmete_kood = K.mõõtmete_kood
    
```

Dimensioon (2)

- Kimballi metoodika järgi dimensioonitabel denormaliseeritakse
 - Dimensioon peab kõiki subjekti kohta käivaid tunnuseid
- Dimensiooni atribuutide valikul arvestatakse:
 - Kodeeritud väärtuste taga on sageli äriloogika
 - nt. kui kaubakoodi esimesed 3 numbrit tähistavad valmistamisriiki, järgmised kolm valmistajatehast, siis luuakse dim. tabelisse vastavad atribuudid.
 - Iga atribuudi väärtus peab olema mõistetav ilma täiendavate vastetabelite
 - nt. S = mees, N = naine, E = ???
 - Andmed dimensioonitabelis on reeglina diskreeditud. Arvuliste atribuutidele saab luua sõltuvalt ärivajadusest intervale.
 - Nt. atribuudi "laste arv" väärtuseid võib jagada vahemikesse {0, 1..2, 3..5, 6+ }

Fakt

Fact

- Fakt tähistab äriprotsessi kohta kogutud huvipakkuvaid andmeid
 - Nt. müügi kohta soovime teada käive ja kasumit
- Üksik kirje faktitabelis tähistab ühte mõõtmist
 - Nt. ühel kuupäeval ühele konkreetsele kliendile müüdud kauba koguseid ja summasid.
- Faktitabeli kirje detailsust nimetatakse granulaarsuseks (grain, granularity)
 - Granulaarsuse määrab dimensioonide arv - mida rohkem, seda detailsem
- Lisaks dimensioonidele saab granulaarsust mõjutada kuupäevadega:
 - Mõõtmine kuupäeva ja kellaaja täpsusega - kasutatakse dimensioone Kuupäev ja Kell.
 - Mõõtmine kuupäeva täpsusega - vastab tüüpiliselt lähtesüsteemi tehingute detailsusele.
 - Kuu kokkuvõtted
 - Pikemate perioodide (kvartalite, aastate) kokkuvõtted.
- Üksiku faktitabeli kirje granulaarsuse määravad ärivajadused:
 - Eelarvete jm. finantsandmete puhul vaadeldakse andmeid aastate või kuude lõikes
 - Müügitehingute dünaamika jälgimiseks on vajaliku kuupäevaline täpsus.
 - Tootmisprotsessi analüüsimiseks võivad olla vajalik tootmismahud vahetuse või veerandtunni täpsusega

Klient	Kaup	Kuupäev	Kogus	Summa
K0462	Q64383	28.10.2011	1	326
T3650	E24664	27.10.2011	2	35
M463	X32464	30.10.2011	5	79

Kontc	Periood	Eelarve	Tegelik
3211	2011-apr	3 217 172	3 345 859
3250	2011-mai	574 337	677 718
4110	2011-apr	2 513 770	2 011 016
4112	2011 mai	321 553	318 337
...

Fakt (2)

- Faktitabel koosneb reeglina ainult välisvõtmetest (koodid) ning mõõdetud väärtustest e. mõõdikutest (*measures*)

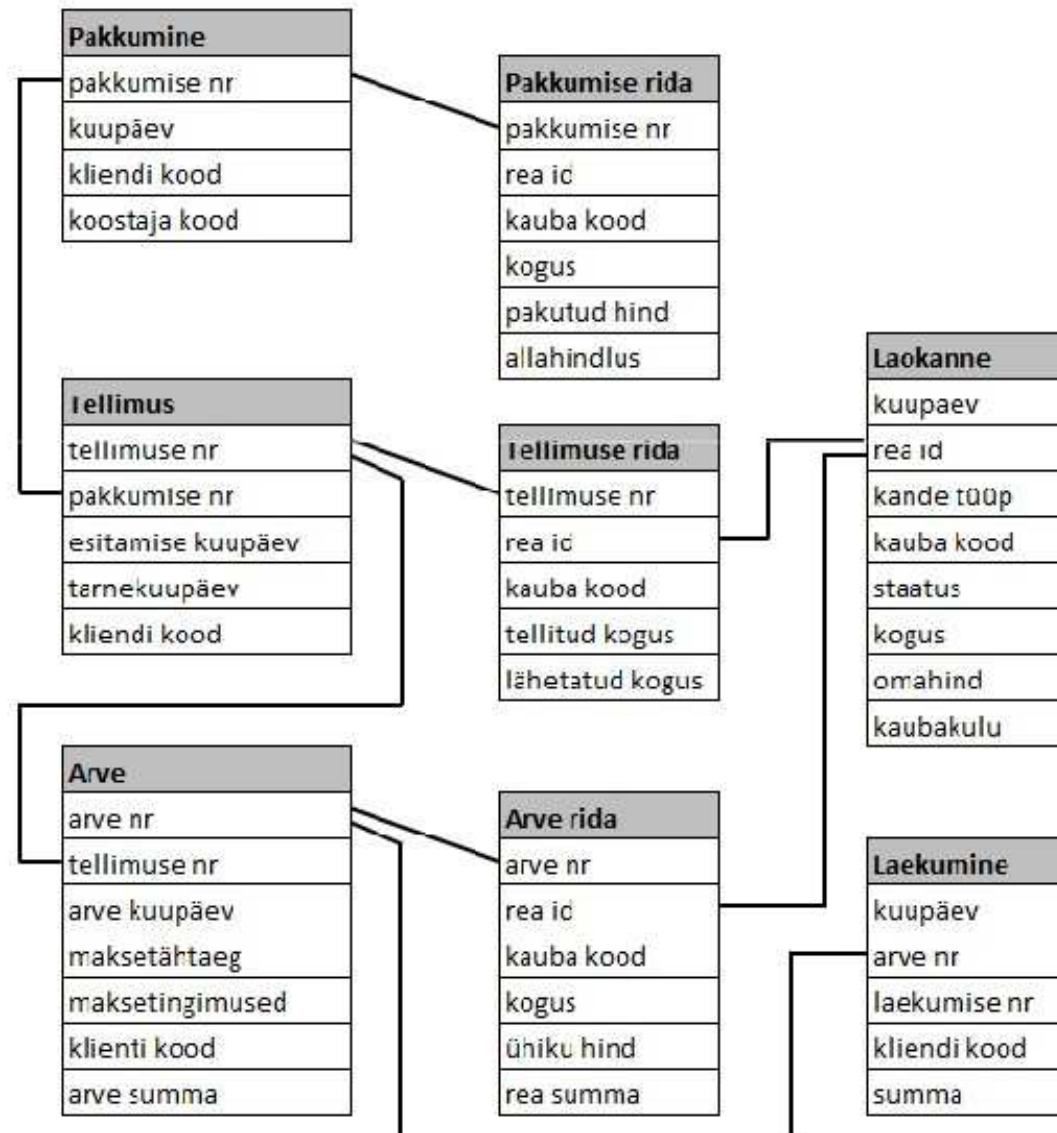
Klient	Kaup	Kuupäev	Kogus	Summa
K0462	Q64383	28.10.2012	1	326
T3650	F24664	27.10.2012	2	35
M463	X32464	30.10.2012	5	79
...

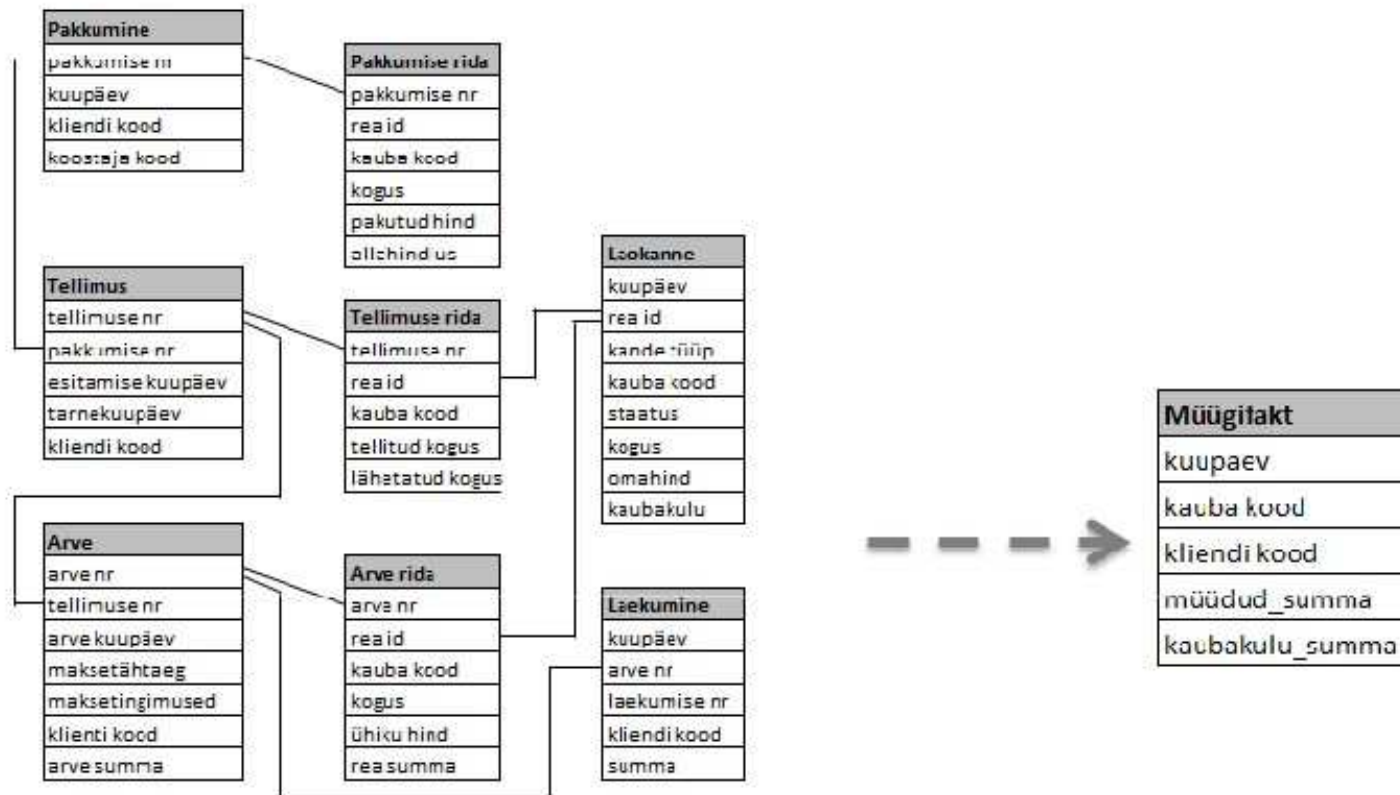
- Mõõdetavad väärtused on arvulised ning neid on kahte liiki:
 - Liidetavad e. summeeritavad väärtused (*additive measures*)
 - nt. kogused, summad
 - Võib kasutada kõiki agregaatfunktsioone: *SUM()*, *AVG()*, *MIN()*, *MAX()* jm
 - Mitteliidetavad väärtused (*non-additive measures*), kuna seda arvutav avaldis ei ole matemaatiliselt korrektne:

$$\frac{A}{B} + \frac{C}{D} \neq \frac{A+C}{B+D}$$

- Näited mitteliidetavatest väärtustest:
 - protsentuaalsed väärtused ja suhtarvud
 - keskmised, miinumumid, maksimumid
 - jäägid mingi perioodi lõpu seisuga (saldod)
- Mõtet omavad ainult *MIN()*, *MAX()*, *FIRST()*, *LAST()*
- Faktitabelis on tavaliselt vähe veerge ja palju ridu

Kuidas tekivad faktid?





```
SELECT P.kuupäev AS pakkumise_kuupäev, T.esitamise_kuupäev, A.arve_kuupäev AS kuupäev
      , PR.kogus AS pakutud_kogus, TR.tellitid_kogus, AR.kogus AS müüdüd_kogus
      , AR.rea_summa AS müüdüd_summa
      , L.kaubakulu AS kaubakulu_summa
```

```
FROM pakkumise_rida PR
```

```
  INNER JOIN pakkumine P ON PR.pakkumise_nr = P.pakkumise_nr
```

```
  INNER JOIN tellimuse_rida TR ON TR.rea_id = PR.rea_id
```

```
  INNER JOIN tellimus T ON TR.tellimuse_nr = T.tellimuse_nr AND T.pakkumise_nr = P.pakkumise_nr
```

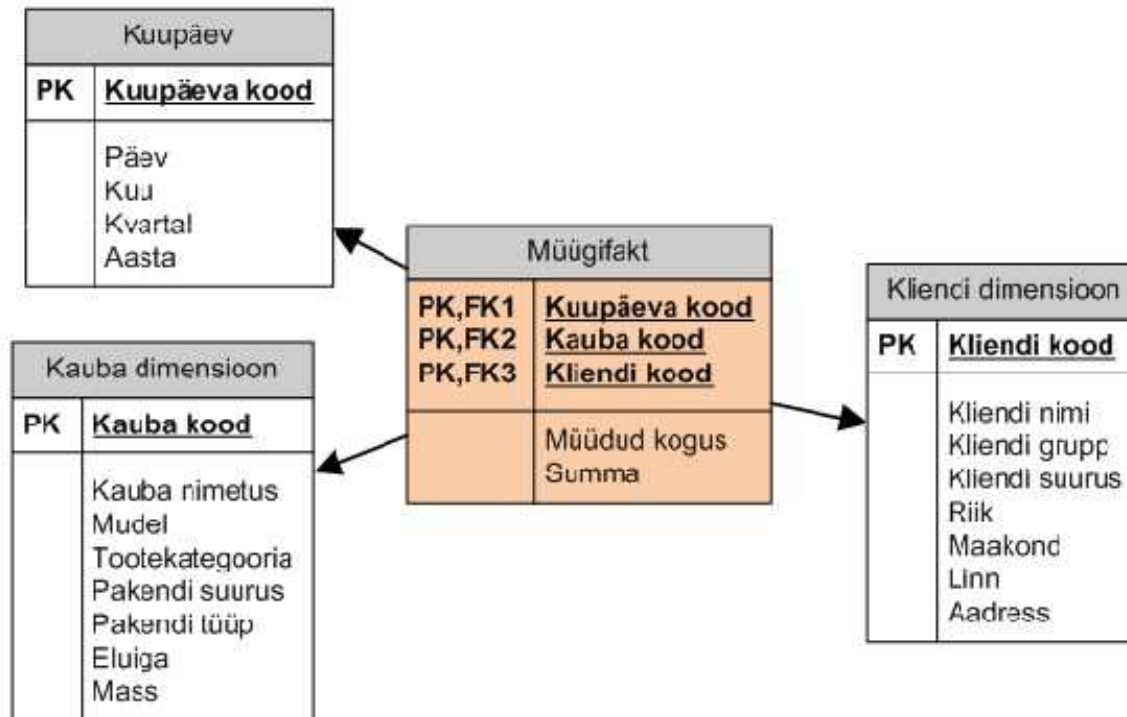
```
  INNER JOIN arve_rida AR ON AR.rea_id = T.rea_id
```

```
  INNER JOIN arve A ON A.arve_nr = AR.arve_nr AND A.tellimuse_nr = T.tellimuse_nr
```

```
  INNER JOIN laokanne L ON L.rea_id = AR.rea_id
```


MultiDimensionaalne andmeMudel

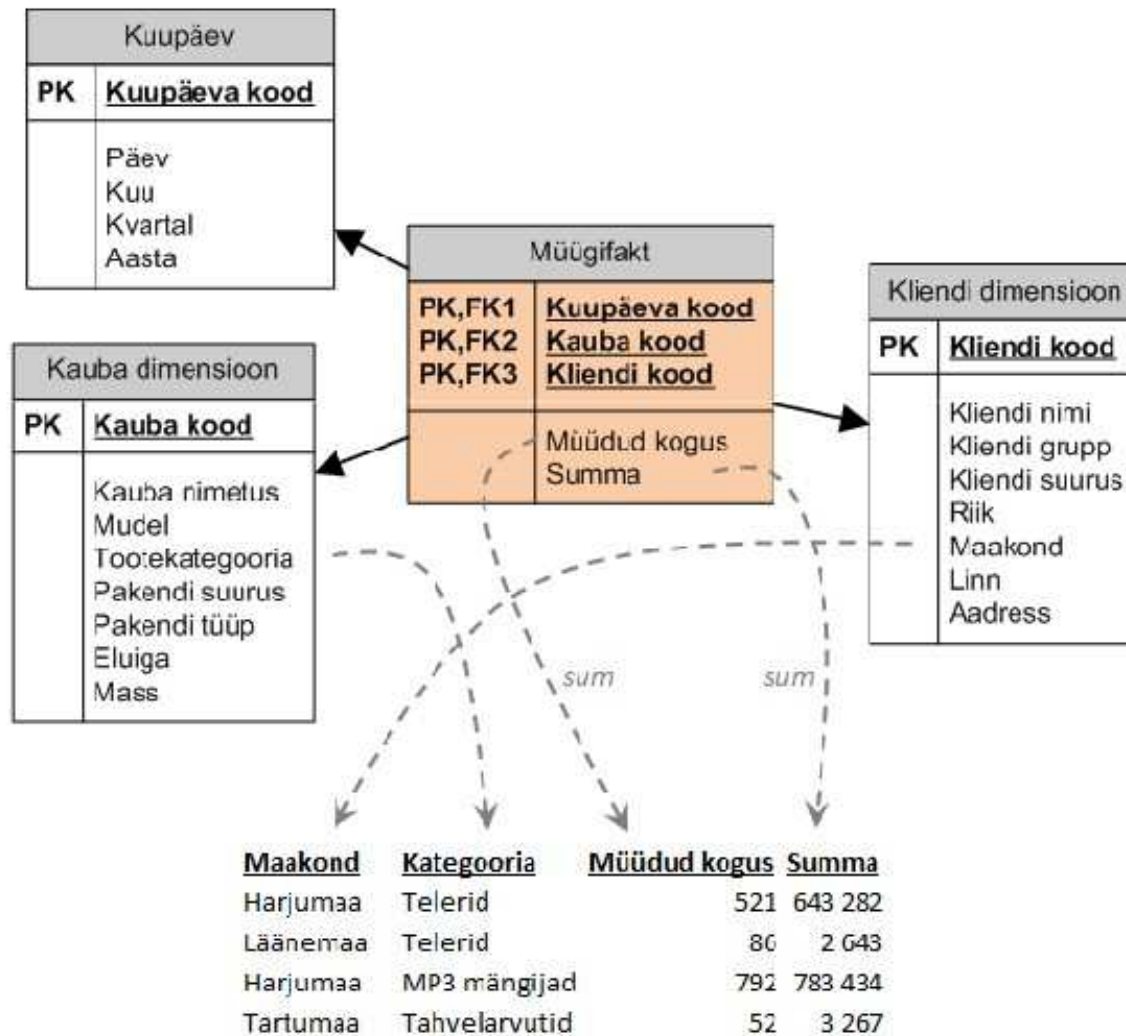
- Koosneb protsesside kohta andmeid talletavatest faktitabelitest ning fakte kirjeldavatest dimensioonidest
- Dimensionaalse mudeli skeem meenutab tähte (*star-schema*)



- Heas dimensionaalses mudelis on iga atribuut faktitabelist ühe ühenduse (sisuliselt JOIN-opeeraator) kaugusel

Päring

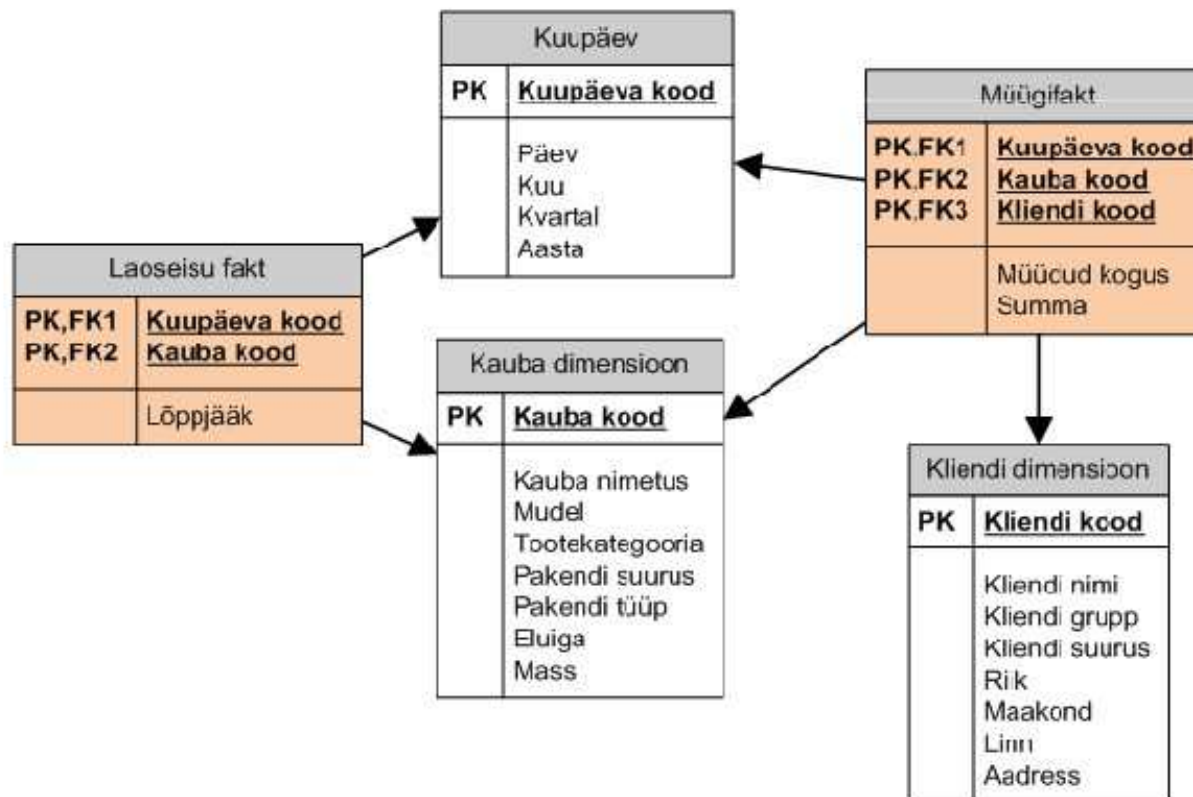
dimensionaalsest andmemudelist



Dimensioonide ühiskasutus

Siiniarhitektuur

- Kui sama subjekt on kasutusel mitmes faktitabelis, siis kõik faktitabelid peavad viitama ühele ja samale dimensioonitabelile
 - Faktide lõikes on subjektid (nt. klient, kaup) ühtlustatud
 - Kõik faktitabelid kasutavad samu välisvõtmeid



Kokkuvõte - kordame mõisted üle

- Decision Support System - otsustust toetav süsteem e. nõustussüsteem e. juhtimisinfosüsteem.
 - DSS on infosüsteem, mis varustab ettevõtte või organisatsiooni otsustajaid neile vajaliku informatsiooniga strateegiliste, operatiivsete juhtimisotsuste, mis on vajalikud täideviivate või planeerimistegevuste läbiviimiseks
- Data warehouse - andmeait (andmeladu)
 - Andmeait on süsteem, mis annab konsolideeritud vaate terve ettevõtte andmetele. Andmed andmeaidas on puhastatud, teisendatud, katalogiseeritud ja süstematiseeritud.
- Data mart - andmevakk
 - Hõlmab ühe valdkonna või äriprotsessi kohta käivaid andmeid, mis võivad sageli olla summeeritud kujul.
 - Andmevakk võib olla osa andmeaidast, kuid süstematiseerimata kogum andmevakkasid ei moodusta automaatselt andmeaita
- Kontrollküsimus: mis on fakt? dimensioon?

Kodune töö

- Läbi lugeda allikast [3] peatükk 6 -
Dimensional Modeling Fundamentals
- E-raamat on loetav TTÜ e-raamatukogus
portal.ttu.ee vahendusel

Viited

- [1] W.H. Inmon - Building the Data Warehouse, 3ed. Wiley, 2005.
- [2] R.Kimball, M.Ross - The Data Warehouse Toolkit: The Complete Guide To Dimensional Modeling, 2ed. Wiley, 2002.
- [3] R.Kimball, M.Ross et. al - The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence. Wiley, 2010.

Loetav TTÜ e-raamatukogus:

<http://site.ebrary.com/lib/ttul/docDetail.action?docID=10369792>