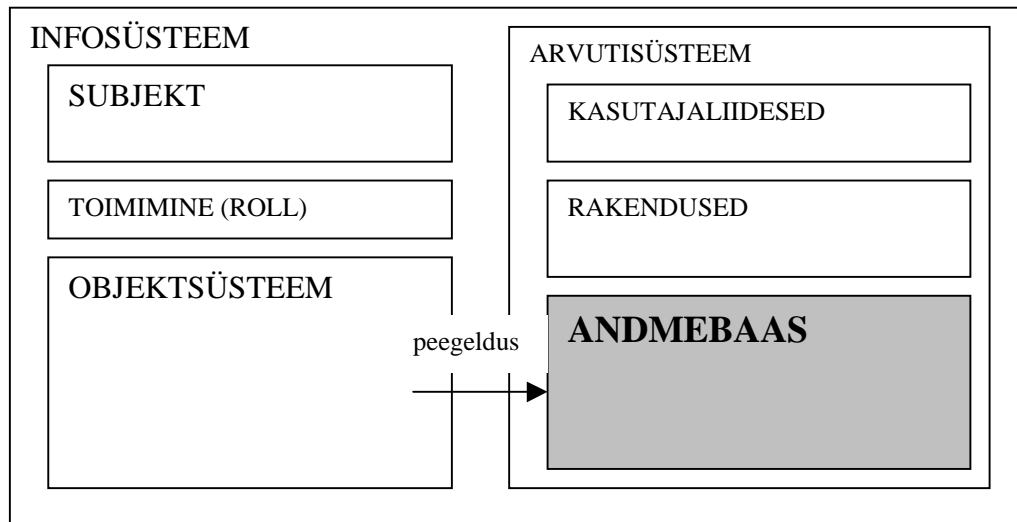


Loeng 1. Andmebaaside põhimõisted

Loengu eesmärgiks on anda ülevaade andmebaaside kohast infosüsteemis, esitada andmebaasi definitsioon, käsitleda tema põhimõisteid ning -omadusi.

Mis on andmebaas ? Tema koht infosüsteemis ?



- Andmebaas on (hästi)organiseeritud andmekogum objektsüsteemi (probleemvaldkonna, subjekti) kohta, mis peegeldab objektsüsteemi seisundit ning selle muutumist ja vastab kõigile korrektselt formuleeritud päringutele objektsüsteemi kohta.
- Andmebaas on infosüsteemi tuum ning arvutisüsteemi komponent, milles organiseeritakse infosüsteemi andmed.
- Andmebaasi põhiülesandeks, -funktsiooniks on genereerida (koostoimes infosüsteemi teiste komponentidega) kvaliteetne, adekvaatne infokeskkond objektsüsteemi seisundi ja selle muutumise kohta infosüsteemi subjektidele (kasutajatele).

Andmebaasi omadused

Andmebaas peab olema:

- Ammendav
- Terviklik
- Mitteliiane
- Struktuurne
- Aktuaalne
- Aktiivne

Andmebaasi elutsükkel

- Objektsüsteemi Analüüs
- Andmebaasi (ja tema rakenduste) projekteerimine (Disain)
- Andmebaasi (ja tema rakenduste) loomine (Ehitamine)
- Andmebaasi (ja tema rakenduste) kasutuselevõtt (Rakendamine)
- Andmebaasi (ja tema rakenduste) hooldamine (Hooldus)

Andmebaaside erinevad liigitused

Relatsioonilised andmebaasid

Navigatsioonilised andmebaasid

Tekstiandmebaasid

Objektiandmebaasid

Lokaalsed andmebaasid

Klient-Server andmebaasid

Hajusandmebaasid

Desktop andmebaasid

SQL andmebaasid

Multidimensionaalsed andmebaasid

Relatsiooniliste andmebaaside põhimõisted.

1.Relatsiooniline andmebaas

- eesmärgiks salvestada, teha kättesaadavaks, hoida informatsiooni.
- seda informatsiooni on vaja hoida mingite objektide kohta (*töötaja, arve, auto* jne.) Nendest nn. infoobjektidest saavadki ühe relatsioonilise baasi andmeobjektid e. tabelid.

Nii hoiame tabelis “TÖÖTAJA” andmeid töötajate kohta.

Käsitletavas kursuses on kõik andmebaasi tabelid organiseeritud relatsioonilise andmearhitektuuri v.mudeli alusel. Esimesena esitas relatsioonilise andmebaasi põhimõisted 1970 a. dr. E. F. Codd. 1970-ndate lõpupoole hakkasid reaalsesse kasutusse jõudma juba ka relatsioonilisele põhimõttele ülesehitatud andmebaasisüsteemid.

Relatsioonilises andmebaasis on andmed organiseeritud teatud põhimõtete järgi, mis seotud terviklikkuse reeglitega (*integrity rules*) .

Andmed relatsioonilises baasis on koondatud tabelitesse, millede hulk kokku moodustabki relatsioonilise andmebaasi. Tarkvara , mis võimaldab selliseid andmebaase luua ja kasutada, nimetatakse (relatsioonilisteks) andmebaasisüsteemideks.

Tabelid koosnevad ridadest ja veergudes. Tabel esitab informatsiooni mingi objekti (õigemini küll ühetüübiliste objektide hulga) kohta - “ISIK”, “ARVE “ oleksid tüüpilised andmetabelite nimed, kuigi on ka tabeleid, millele enam nii lihtne kohast pealkirja leida ei olegi.

veerg esitab selle andmeobjekt objekti mingit omadust e. atribuuti. Atribuut tähistab ühte tüüpi infot, mis on salvestatud selle objekti kohta. Rida tähistab ühte kirjet, st. ühte konkreetset objekti.

Eristame - abstraktne objekt (objektiklass, tabel)
-konkreetne objekt (kirje, konkreetne rida tabelis)

Näide

TÖÖTAJA

NIMI	ISIKUKOOD	SÜNNIAASTA	AADDRESS
Mati Jõgi	3456723437	1956	Tallinn, Ehitajate 23
Andres Mets	5688934895	1972	Rapla, Jõe tn.18
Tiiu Lepik	3456789899	1970	Tallinn, Pärnu mnt.38

veerg (atribuut)

rida (üks kirje, üks konkreetne töötaja - Mati, Andres jne.)

tabel - "TÖÖTAJA"

atribuudid - "NIMI", "ISIKUKOOD", "SÜNNIAASTA", "AADDRESS"

tabeli üks kirje - [" Andres Mets ", 5688934895 , 1972, "Rapla, Jõe tn. 18"]

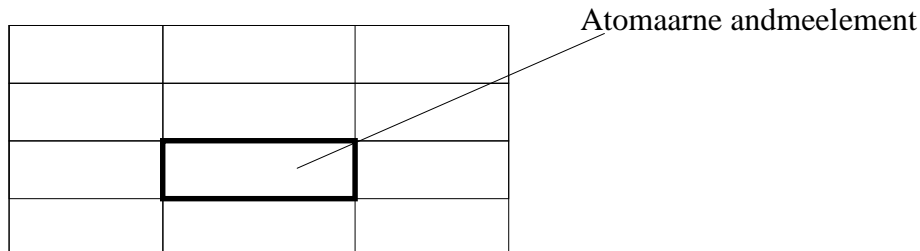
abstraktne objekt , objektiklass - "TÖÖTAJA", st. töötajate klass

konkreetne objekt , konkreetne selles tabelis salvestatav töötaja - (n. Andres Mets)

2. Relatsioonilise andmebaasi põhiomadused

- Tabelid , tabelite omadused

- igas veerus on sama tüüpi andmed (samad andmetüübid, samad objekti omadused)
- iga rida (kirje) peab olema sisu (väärtuse)poolest unikaalne, st. eristatav tabeli teiste kirjete hulgast.
- relatsioonilises mudelis ridade(tabeli kirjete) ja veergude (tabeli atribuutide) järjekord tabelis andmete kasutamisel ei ole tähenduslik, st. ükskõik , mis järjekorras nad on, sellest ei muutu tabelis salvestatud informatsiooni sisu.
- iga veeru-rea sisu on nõ. *atomaarne andmeelement* st. kogu süsteemis salvestatavad andmed on mingis mõttes lõigatud, jagatud nn. atomaarseteks andmehulkadeks, mida enam jagada ei saa ja siis on nad pandud relatsioonilisse andmebaasi.



Andmete atomaarsus, jaotamine relatsioonilisse andmebaasi.

Andmetüübid - tabeli iga atribuut kuulub mingisse kindlasse andmetüüpi, mis näitab, mis tüüpi andmeid antud veerus hoitakse. Põhilised andmetüübid, mida võib kohata enamikes andmebaasisüsteemides, on :

 CHARACTER - tekstitüüpi andmete salvestamiseks
 [“Andres Mets”]

 INTEGER - täisarvtüüpi andmete salvestamiseks
 [5688934895]

 FLOAT - reaalarvutüüpi muutujate salvestamiseks [3,1456]
 DATE - kuupäevade salvestamiseks [“10/10/94”]

Iga andmebaasisüsteem pakub lisaks ülaltoodud põhilistele andmetüüpidele veel enamasti lisaandmetüüpe, mis erinevates andmebaasisüsteemides on erinevad (n. MONEY raha salvestamiseks, VARCHAR muutuva pikkusega tekstiväljade salvestamiseks jne.) Samuti võivad erinevates andmebaasisüsteemides olla samade andmetüüpide nimetused olla erinevad (NUMBER, NUMERIC jne.)

Paljude andmetüüpide korral on võimalik ette andma ka andmeväljade pikkust:

 NUMBER(10) võimaldab salvestada kuni kümnekohalisi arve,

 VARCHAR(8) kuni kaheksakohalisi tekstivälju

- Andmestruktuuride normaliseeritus

Normaliseerimine - andmestruktuuride organiseerimine sel viisil, et neist struktuuridest kaoksid korduvad andmeelemendid, et ei oleks dubleeritud andmeid süsteemis, Andmeliiasuse kõrvaldamine.

Näide selle kohta

töötaja_nr	nimi	aadress	ameti_nr	ameti_nimetus
34	ANTS	Jõe tn.3	2	keevitaja
34	ANTS	Jõe tn.3	5	autojuht
34	ANTS	Jõe tn.3	6	mehhanisaator

“TÖÖTAJA” tabelnormaliseerimata kujul, töötaja andmed dubleeruvad, kuna “ameti” näol on tegemist korduvate andmelementidega. Põhjus : üks-mitu suhted töötaja ja tema ametite vahel - ühele töötajale vastab mitu ametit

töötaja_nr	nimi	aadress
34	ANTS	Jõe tn.3

töötaja_nr	ameti_nr	ameti_nimetus
34	2	keevitaja
34	5	autojuht
34	6	mehhanisaator

Samad andmed normaliseeritud kujul, kahe tabelina, andmete liiasus likvideeritud (va. töötaja_number, mis loob relatsiooni andmetabelite vahel). See on *esimene normaalkuju* ja sellisel viisil järgitakse esimese normaalkuju reeglit. Miks selle reegli järgimine vajalik - põhjus pole mitte ainult andmebaasi mahu raiskamises, vaid eelkõige andmete õigsuse tagamine andmebaasis.

Normaalkujusid on vähemalt viis, kolme esimese täpsemad definitsioonid tulevad hiljem.

- Relatsioonid, põhiomadused ja eesmärk.

Relatsioonilises andmemudelis on nii, et seosed andmeobjektide (tabelite) vahel on väljendatud ja loodud vastavate andmete salvestamisel vastavatesse tabelitesse, need andmed loovadki relatsiooni e. seose erinevates tabelites olevate kirjete vahel. Relatsioonide loomiseks võetakse, tõmmatakse ühe tabeli ühe (või ka mitme) veeru andmed teise tabelisse salvestamiseks. Tekkinud kahe erineva tabeli ühesuguse sisuga nn. “seosväljad” loovadki relatsiooni nende tabelite vahel.

Relatsioonide mõte - nad loovad meile lisainformatsiooni, relatsioonid ongi suure osa info tekitajad süsteemis - mis on millega seotud . Tabelid, mis on seotud relatsioonidega, annavad lõppkokkuvõttes tunduvalt rohkem infot, kui need tabelid eraldi. Vt eelmist näidet , relatsiooni seal loob mõlemas tabelis (“TÖÖTAJA, “AMET”) salvestatud töötaja_nr, mis annab seose töötajate ja nende ametite vahel.

- Võtmed

Nagu oli öeldud relatsioonilise andmebaasi põhidefinitsioonis, peavad tabeli kirjed, read olema väärtuse mõttes unikaalsed, st nendes ridades peab olema midagi, mis neid üksteisest eristab. Tavaliselt kindlustab seda ka andmebaasisüsteemi tarkvara, mis ei lase salvestada identseid kirjeid või lisab siis ise midagi eristavat juurde.

Seega peab olema tabelis 1 või rohkem veergu , mille väärtused on iga juhul iga kirje puhul teistest erinevad, st. unikaalsed. Seda veergu, või veergude kogumit nimetatakse *primaarvõtmeks*

- primaarvõtme väärtus peab igas reas olema erinev teiste primaarvõtmete väärtustest selles tabelis (st. ta peab eristama kirjeid)
- kui primaarvõtme osiseid muuta või kustutada, võib see rikkuda unikaalsuse nõuet.

Mõned andmebaasisüsteemid ei nõua primaarvõtmete defineerimist otseselt, kuid nõ. “sisemiselt” teevad neid ikkagi. Primaarvõtmeid kasutatakse tabelite sidestamisel relatsioonide loomiseks ,vt. Relatsioonide juttu.

KOKKUVÕTTEKS: relatsiooniline andmebaas kujutab endast teatud viisil organiseeritud ja omavahel seotud andmetabelite kogumit

- iga tabel esitab mingit objekti tüüpi, mille kohta on vaja infot salvestada
- tabeli veerg on üks objekti tüübi omadus e. atribuut.
- tabeli ritta on siis salvestatud antud objektitüübi konkreetne nn. esinemisjuht, konkreetne isik, arve, auto, amet jne.
- Andmetabelil on järgnevad karakteristikud ja omadused :
 - iga veerg-ritta tehtud sissekanne kujutab endast nn atomaarset, jagamatut andmeelementi.
- andmete dubleerimine ja liiasus pole lubatud, andmestruktuurid peavad olema normaliseeritud
- tabelite veerunimed peavad olema unikaalsed
- iga tabeli rida (kirje) peab olema unikaalne
- ridade-veergude järjekord tabelis pole oluline ja ei muuda tabeli sisu.
- primaarvõtmete olemasolu tabelis.
- kõik suhted e. relatsioonid tabelite vahel on väljendatud andmetes, mis on salvestatud nendesse tabelitesse.