



Sissejuhatus

Sisejuhatatus

- Kaasaegsete andmete modelleerimise ja andmebaasisüsteemide eelne periood
- Kaasaaegse andmete modelleerimise eelkäijaks võib lugeda erinevaid ankeetide ja kartoteekide koostamise metoodikaid, samas kui kaasaaegsete andmebaasisüsteemide eelkäijaks on ankeetide kogumid ja kartoteegid
- Nende teooriate ja rakenduste suurimad arendajad ja kasutajad on olnud luureorganid ja kriminalistika asutused

Sisejuhatatus

- 1881 aastal alustati Prantsusmaal Alphonse Bertillioni juhtimisel kurjategijate registreerimist ja kartoteekimist antropomeetrilisel meetodil, kus andmete süstematiseerimise aluseks oli kurjategijate kehaosade mõõtmed.
- Koos registri loomise ja pidamisega käis pidevalt andmete süstematiseerimise ja andmeotsingu metoodikate arendustöö.

Sisejuhatatus

- Kui registri arendamise algusaegadel piirduti inimese identifitseerimiseks 14 kehaosa mõõdetega siis aja jooksul lisandusid sinna foto, isiku struktureeritud kirjeldus ja lõpuks isegi mõne etteantud sõrmede jäljed.
- Andmete keerukamaks muutumisel muutusid keerukamaks ka otsingumeetodid – lisaks mõõtmistulemustele pidi olema lõpuks võimalik leida kartoteegist isiku kaarti ka kirjelduse üksikute elementide ja sõrmejälgede järgi.
- 1914-ks aastaks, kui antropomeetria asendati Prantsusmaal daktüloskoopiaga ja kogu kartoteegi struktuur läks ümberkorraldamisele, oli kartoteegis sadu tuhandeid kaarte.

Sisejuhatatus

- USA-s loodi 1930. aastal FBI juurde isikusamastamise register, mis põhines daktüloskoopial. Selle registri loomise juures on tähelepanuväärne see, et registrit ei loodud tühjalt kohalt, vaid sinna koondati alates 1911-st aastast osariikide ja linnade politseijaoskondades olevad kartoteegid.
- Ületoomise ja liitmise käigus tehti ära suur töö väikestes registrites olevate andmete struktuuride ühtlustamiseks, andmete süstematiseerimiseks ja suurt andmehulka toetava otsingusüsteemi väljatöötamiseks.

Sisejuhatatus

- Kui registri loomise algusaegadel toimus kogu töö käsitsi, siis aastaks 1956 a oli maht juba ca 14 miljoni kaardini (ca 140 miljonit sõrmejälge), siis toetas kaartide kartoteeki paigutamist ja sealt nende leidmist juba arvutisüsteem.
- Kuna arvutid ei olnud veel oma võimalustelt sellised, mis võimaldaks kogu informatsiooni salvestada elektrooniliselt, siis oli andmete adresseerimissüsteem lahutatud andmetest enestest – arvudeks kodeeritud tunnuste alusel leidis arvuti kartoteegikaardi aadressi kartoteegis.

Sisejuhatatus

- Toodud näited ei kirjelda muidugi esimesi registreid või kartoteeke, kuid need näitavad ilmekalt seda, milliste rakendusteni oli maailm jõudnud enne seda kui hakkasid arenema kaasaaegne andmete modelleerimine ja andmebaasisüsteemid.

Tekkimise eeldused

- Elektrooniliste andmebaaside (esialgu küll kartoteekide) loomisele andsid peamise tõuke kolm asja (antud järjekorras) – kaasaegse struktuuriga arvuti loomine, aine magnetilistel omadustel baseeruva püsimalu loomine ja andmete lugemise-kirjutamise otsepöördus (mitte-järjestik) süsteemide loomine.
- Siin ei saa muidugi tähtsa katalüsaatorina jätta mainimata seda, et selleks ajaks oli loodud hulk strateegilise tähtsusega kartoteeke, mille pidamine vanal tehnilisel ja struktuursel platvormil hakkas üle jõu käima.

Tekkimise eeldused

- Kaasaaegse arvuti esimese põlvkonna (1945-1956) tekkimisel oli suur osakaal II maailmasõjal, kus riigid püüdsid arvutite abil saavutada strateegilist üleolekut – sõja ajal tehtud pingutused kanaliseerusid vahetult pärast sõda uue põlvkonna arvutite loomisega, milles rakendatud põhimõtted määrasid arvutite arendamise trendid järgmiseks neljakümneks aastaks (EDVAC - University of Pennsylvania 1945, EDSAC - Cambridge University 1949, UNIVAC I - Remington Rand, 1951).

Tekkimise eeldused

- Peamisi põhimõttelisi muutusi oli kaks.
- Esiteks kirjeldati arvuti arhitektuuris keskjuhtseadme (protsessori) mõiste, mis võimaldas arvutit juhtida läbi ühe andmevoo.
- Teiseks hakati nii täidetavat programmi kui ka programmi juhtimiseks või töötlemiseks vajalikke andmeid hoidma samas mälus. See kõik lõi aluse kaubanduslike arvutite tootmiseks.

Tekkimise eeldused

- 1945. aastal loodi uus andmekandja, magnetlint, mis vähehaaval hakkas välja vahetama perfokaarte ja perfolinte.
- See oli esimene andmekandja, mis võimaldas andmete otsimist. Siiski ei olnud see otsingumehhanism veel eriti täiuslik, kuna võimalik oli ainult järjestik-otsing.
- Suure tähtsusega oli aga see, et “ruumiühikusse salvestatud andmete hulk” kasvas mõõtnatult ja andmete fragmenteerimisega paljudele paralleelselt töös olevatele lintidele suudeti tagada ka juba piisav operatiivsus andmete otsimisel.

Tekkimise eeldused

- Esimeste kaasaegsete andmebaasisüsteemide loomiseni oli aega veel umbes paar aastat, kui William. C. McGee 1959. aastal publitseeris oma artikli “Generalization: Key to Successful Electronic Data Processing” ajakirjas Journal of the ACM (Volume 6, Number 1, January 1959 lk. 1-23, ACM – Association for Computing Machinery).
- Artiklis ei anta küll veel konkreetseid andmete üldistamise põhimõtteid vaid piirduakse üldkontseptuaalse filosoofiaga, kuid see on esimene suunda näitav teeviit kaasaegse andmete modelleerimise põhimõtete poole.

Tekkimise eeldused

- Samal aastal võtab IBM kasutusele oma magnetkettasüsteemi Ramac (Random Access Method of Accounting and Control) mudel 305, mis on maailma esimene magnetkettasüsteem ja mis koosneb 50-st umbes 60 cm diameetriga kettast, millede mõlemale küljele saab salvestada informatsiooni.
- Informatsiooni salvestamise tiheduseks on 2000 bitti ruuttollil kogumahuga 5MB. Olulisimaks murranguks on siin siiski andmete otsepöördus lugemis/kirjutusrežiimi esmakordne rakendamine ja andmete kirjutamise/lugemise suur kiirus.

Tekkimine

- Aastal 1961 töötati korporatsioonis General Electric Co. välja andmebaasi juhtimissüsteem IDS (Integrated Data Store), mida loetakse esimeseks elektrooniliseks andmebaasi juhtimise süsteemiks.
- Projekti juht oli Charles Bachman. Siin ei saa muidugi rääkida veel andmebaasi juhtimissüsteemist tänapäevases tähenduses, kuna enamik andmebaasi funktsioone oli käsitsi kodeeritud, andmebaasina oli käsitletav ainult üks fail ja ta töötas ainult General Electric Co. arvutitel ja lahendas ainult selle firma konkreetseid vajadusi.

Tekkimine

- Sellest sai tõuke CODASYL (Conference on Data Systems Languages) grupi kokkukutsumine, mis koosnes vabatahtlikest ja mille eesmärgiks oli andmesüsteemide efektiivsema analüüsi, disaini ja rakendamise vahendite ja metoodikate väljatöötamine.
- Grupp moodustati 1959 aastal ja töötas kuni 1985-nda aastani. Grupi peamiseks ülesandeks määratleti standardse, erinevatel arvutitel kasutatava programmeerimiskeele loomine. Selleks programmeerimiskeeleks sai COBOL ja selle raames formuleeriti ka võrk-andmemudeli põhilised kontseptsioonid.

Tekkimine

- 1968 aastal tuli IBM välja oma IMS (Information Management System) kontseptsiooniga, mis formuleeris hierarhiliste andmemudelite põhialused.
- Seejärel kohe (1969) sama mudeli täiendusega (IDM DB/DC), mis kirjeldas meetodit võrk-vaadete ehitamiseks hierarhilisele andmemudelile. Mõlemad lahendused olid mõeldud kasutamiseks IBM SYSTEM/360 mainframe'del.
- Kuni siiani olid andmebaasid kõik ühe protsessi poolt kasutatavad, siis eelmise sajandi 70-ndate aastate lõpus lõi IBM koos American Airlines'ga süsteemi SABRE, kus läbi kommunikatsioonivõrgu said andmetele korraga ligi juba paljud kasutajad.

Tekkimine

- Vaatamata suhteliselt tormilisele erinevate andmebaaside loomisele eelmise sajandi 70-ndate aastate lõpus, ei olnud ikka veel tekkinud kommertsiaalsetel alustel müüdavat, ristvarast eraldiseisvat andmebaasisüsteemi.
- Sellele aluse panemiseks oli jällegi suur teene firmal IBM. IBM uurija Edgar F. Codd pakkus 1970 aastal välja relatsioonilise andmebaasimudeli, kus andmeid säilitatakse tabelites, mille vahele ehitatakse relatsioonid. IMS mudelit täiendati relatsioonilise andmemudeli printsiipidega ja selle alusel arendati välja andmebaasisüsteem SYSTEM/R, mida müüdi koos IBM mainframe'dega kuni 1980-nda aastani.

Tekkimine

- IBM süsteemi SYSTEM/R kohta avaldatud informatsiooni võtsid oma uurimis- ja arendustöodes aluseks California Ülikooli teadlased Michael Stonebraker ja Eugene Wong ning lõid nende arenduste tulemusena oma andmebaasisüsteemi, millele andsid nimeks Ingres ja millel olid kõik kommertsialiseerimiseks vajalikud olulised omadused.
- See produkt kommertsialiseeriti lõpuks firmade Oracle Corp. ja Ingres Corp. poolt.

Tekkimine

- Eelmise sajandi 70-ndate keskel toimus veel üks andmete modelleerimise tähtsündmusi, millest ei saa mööda vaadata.
- 1976. aastal esitles Peter P. Chen olemi-suhte diagrammi, mis on kaasaajal üks kõige levinumaid andmete modelleerimise metoodikaid.
- Huvitav on siinkohal Chen'i väide, et selle teooria kirjapanemiseks oli ta kogu elu, kuni selle hetkeni kogu aeg õigel ajal õiges kohas ja tal jäi ainult elu enese poolt ettelükatud metoodika kirja panna.

Tekkimine

- Aastast 1979 tasub ära märkida veel sündmus, mis märkis ühe tõusva produkti ja trendi sünni. 70. aastate jooksul töötati välja mitmeid erinevaid päringukeeli – SQUARE, SEQUEL, QBE, QEL jne. 1979. aastal lõi Oracle turule esimese kommertsiaalse andmebaasisüsteemi, mis kasutas andmemanipuleerimiskeelena SQL-keelt (Oracle Corp. teisend standardist SEQUEL). Tõusvaks produktiks sai Oracle DBMS ja tõusvaks trendiks SQL-keel.

Arendused 1980-ndatel

- 60-ndate aastate lõpust hakkas arenema veel üks süsteemide grupp, mis tänapäeval on suurte andmebaaside kasutamise lahutamise koostisosa.
- Esialgu nimetati neid otsuse toetussüsteemideks (DSS – Decision Supporting System) ja nende esmane eesmärk oli andmekäsitluse lihtsustamine ja parem kasutamine otsuste tegemise toetamisel.
- Sisuliselt oli tegemist andmete käsitlemise analüütiliste vahendite arendustega. Selliseid alamsüsteeme loodi küll kogu eelmise sajandi 70-ndate aastate jooksul kuid esimese kommertsiaalse lahenduse ni jõuti alles 1970 aastal. Selleks oli süsteem EXPRESS.

Arendused 1980-ndatel

- Eelmise sajandi 80-ndate aastate alguseks oli põhiline platvorm andmete modelleerimise ja andmebaasisüsteemide arenemiseks loodud ja tundus, et mingisuguseid erilisi lööke ei edasi ega tagasi ei tohiks tulla.
- Selle situatsiooni lõi aga segamine personaalarvutite turuletulek. Üpris pea loodi esimene personaalarvutile mõeldud relatsiooniline andmebaasisüsteem DBase. Kiiresti järgnesid DBase II, Paradox, Fox, FoxPro, DBase III, Dbase IV jne.
- See muutis olukorda pöördeliselt – andmebaasid said kättesaadavaks suurele kasutajate hulgale ja andmete modelleerimise meetodikad ja vahendid vajalikuks paljudele. Arendama asuti järjest uusi ja mugavamaid kasutajaliideseid.

Arendused 1980-ndatel

- Mõne aja pärast lisandusid lokaalvõrkude loomise vahendid – riistvara ja tarkvara. Oma hüppe tegid kaasa ka andmebaasisüsteemid – kõigile elujõulistele andmebaasisüsteemidele lisati andmete mitme kasutaja poolt samaaegse kasutamise omadused.
- See ei osutunudki aga nii lihtsaks kui esialgu tundus – hakkasid tekkima konfliktsituatsioonid andmete samaaegsete kasutajate vahel ja see tingis andmete ühiskasutuse ja lukustusteoriate arenemise.

Arendused 1980-ndatel

- 1981 aastal tuli James Martin juhtimisel töötanud uurimissgrupp (JMA – James Martin Associates) välja infosüsteemide loomise metoodika IE (Information Engineering) esimese versiooniga.
- Lisaks J. Martinile kuulusid gruppi veel Clive Finkelstein ja Tong Carter. IE esimeses versioonis keskenduti andmeanalüüsile (infooloogilised skeemid, normaliseerimine, andmetele juurdepääsu-analüüs) ja andmevoo diagrammidele. Aasta hiljem tuldi välja tervikliku IE metoodikaga.
- Selles metoodikas sisaldas ka andmete seos-maatriks, mis sellest tehnoloogiast lahutatuna on ka iseseisvat rakendust leidnud. 1984 aastal liitus grupiga Ian MacDonald ja töötati välja metoodikad andmeanalüüsi, andmevoogude diagrammide ja objekt-modelleerimise ühendamiseks. Tehti ka esimesed arendused analüüsi ja modelleerimist toetava tarkvara loomiseks.

Arendused 1980-ndatel

- 1985-ndal aastal publitseeriti SQL-keele esialgne standard – keel, mille Oracle Corp. oli kasutusele võtnud kui oma andmebaasisüsteemi andmemanipuleerimiskeele.
- Vahepeal oli läbi teinud standardiseerimise protsessi ja valmis kasutamiseks suuremas hulgas produktides.
- Praeguseks on toona kinnitatud standardit mitmeid kordi muudetud ja ANSI klassifikatsiooni järgi on viimane standard SQL:2019, mida arvestavad oma uutes versioonides kõik tuntumad andmebaasisüsteemid.

Arendused 1980-ndatel

- 1985 aasta oli nii mõneski mõttes huvitav aasta. Lisaks SQL-keele esmase standardi esitlemisele rakendati samal aastal ka esimene äri-intelligentsi (business intelligence) süsteem.
- Firma Metaphor Computer Systems Inc. valmistas Procter & Gamble Co. jaoks süsteemi, mis ühendas analüüsiks müügiinformatsiooni ja turu jälgimise informatsiooni. Samal aastal alustas Pilot Software Inc. süsteemi Command Center müüki – turule oli tulnud esimene Client/Server arhitektuuriga süsteem.

Arendused 1980-ndatel

- Seoses sellega hakkasid ka kohe arenema eraldatud (distributed) struktuuriga andmebaaside teooriad ja õigepea ka andmete replikeerimist käsitlevad teooriad ja toetavad süsteemid. Paraku ei ole lõpuni korrektselt toimivate ja kõiki kasutamisevajadusi rahuldavate replikeerimissüsteemide loomiseni jõutud veel tänasel päevalgi.
- Samuti samal aastal muudeti Ingres projekt Postgre projektiks, eesmärgiga luua objekt-orienteeritud andmebaasisüsteem. Siiski ei saanud Postgre mitte esimeseks objektorienteeritud andmebaasisüsteemiks. Selle au näppas ära 1986 aastal firma Graphael Inc., kui ta tuli turule andmebaasisüsteemiga Gbase.

Arendused 1980-ndatel

- Aastal 1988 defineerisid firma IBM uurijad Barry Devlin ja Paul Murphy mõiste andmeladu (information warehouse) ja sellest ajast alates algas andmeladude ehitamine.
- Aastal 1991, avaldas W. H. Inmon raamatu “Building the Data Warehouse”. Sellest alates arvatakse andmeladude praktilise teoreetilise arengu alguseks. Selles raamatus esitati peamised printsiibid andmelattu andmete kollektioneerimise, andmete selekteerimise ja valimite koostamise tehnoloogiate kohta.

Arendused 1980-ndatel

- 80-ndate aastate trendidest on mainimata veel ekspertsüsteemid – teadmiste süsteemid, mis baseeruvad tehisintellektil (artificial intelligence) ja on suunatud vajaliku informatsiooni leidmisele, süstematiseerimisele ja analüüsile suhteliselt kitsas ainevallas.
- Ekspertsüsteemide loomisel on kaks sihifunktsiooni – inimese intelligentsi uurimine ning luues teadmiste baase ja nende käsitlemise mehhanisme ning praktiliste rakenduste loomine erinevate valdkondade töötajate teabevajaduste rahuldamiseks.
- Andmebaasisüsteemide areng ja seega ka informatsiooni talletamise ning analüüsimise vahendite kasutatavuse paranemisega tekkis soodne pinnas seda tüüpi teooriate arenemiseks ja rakenduste loomiseks.

Arendused 1980-ndatel

- 80-ndate teisel poolel tuleb turule uus liik andmete modelleerimist toetavaid süsteeme - CASE-süsteemid. Siin on peamiselt kolm põhjust, miks see just sellel ajal juhtub.
- Esiteks on modelleeritavad andmestruktuurid läinud nii suureks, et nende “käsitsi modelleerimine” on muutunud pea võimatuks.
- Teiseks, tunduvalt paremaks muutunud arendusvahenditele, infosüsteemide dünaamika ja andmenudeleid on vaja pidevalt ringi teha, samas tagades sünkroonselt adekvaatse dokumentatsiooni üha suuremaks muutuvates arendustiimides.
- Kolmas ja mitte vähe oluline põhjus on see, et graafilised info ehituse tehnikad on arenenud piisavalt kaugele selleks, et luua graafilisel andmestruktuuride käsitlelusel põhinevaid arendusvahendeid.

Arendused 1980-ndatel

- Esimesed andmete modelleerimise CASE-süsteemid luuakse traditsioonilistesse graafilisse töökeskkonda, milleks on Macintosh oma MacOS operatsioonisüsteemiga.
- Enamikus toetavad nad JMA grupi poolt loodud IE metoodikat. Üks tuntumaid selliseid vahendeid oli CASE-süsteem DEFT, mis võimaldas luua ERD-mudeleid (Entity Relationship Model) ja DFD-mudeleid (Data Flow Diagram), kommenteerida neid erinevatel tasemetel ja trükkida välja dokumentatsiooni.
- Üsna pea, koos MS Windows' operatsioonisüsteemi arenguga tulid vastavad arendused ka PC maailma. Tuntumateks kaubamärkideks olid ERWin, PBWin (Logic Works), System Architect (Popkin Software) ja SilverRun (Magna Solutions). Nende turule tulemise hetkel oli kõikidel andmete modelleerimise CASE-süsteemidel oluliselt paranenud dokumentatsiooni genereerimise võimalused. Lisaks sellele oli lisandunud andmebaaside automaatgenereerimise ja muudatuste sünkroniseerimise omadused.

Arendused 1990-ndatel

- 1990-ndate aastate arengud on kõik suuremal või vähemal määral seotud interneti, kommunikatsiooni, töökindluse ja multimeediaga.
- Interneti arenemine pööras tagasi mõned leheküljed andmekäsitluse ajaloos. SQL-päringukeele kõrvale tõusis taas kirje-taseme pöördus (RLA – row level access), kui SQL-keelset tunduvalt ebamugavam kuid samal ajal tunduvalt kiirem andmekäsitluse meetod.
- See omakorda on upitanud tõusuteele objektorienteeritud andmebaasid, mille kasutus kuidagi edeneda ei tahtnud, kuna siin on meetodite taha vastavalt vajadusele võimalik maskida just kasutuskeskkonna ja viisi poolt määratud pöördusmeetodid. Kuigi need kasvud ei ole sellised mida tegelikult loota võiks.

Arendused 1990-ndatel

- Kasutajate hulga ja töödeldava informatsiooni mahu pidevast kasvust tingituna on oluliseks suunaks ka mitme protsessoriga arvutite tulek turule ja sellest tingituna andmebaasisüsteemide arenemine kasutamaks seda jõudlust. See nõuab omakorda täiesti uute optimeerimismetoodikate väljatöötamist ja rakendamist.
- Kogu 90-ndaid kattev suur trend ongi SQL päringute uute optimeerimismeetodite väljatöötamine ja rakendamine.
- Kommunikatsioonitehnika arenemisest tulenevalt kasvab 90-ndatel mitmete erinevate andmebaaside ja isegi erinevate andmebaasisüsteemide abil lahendatud andmebaaside kooskasutused. Arenevad replikeerimistehnikad, andmete distantssvarundus, süsteemide distantssdiagnostika jms.

Arendused 1990-ndatel

- Kuni siiani võib tunduda, et eelmise sajandi 90-ndad aastad ei andnudki andmete modelleerimise ja andmebaasisüsteemide arengusse midagi kardinaalselt uut, tegeldi vaid juba olemasolevate teooriate ja tarkvarakomponentide täiustamisega.
- Tegelikult tuli juurde vähemalt kolm uut asja – UML-süsteemid (Unified Modeling Language), spetsiaalsed andmebaasisüsteemid andmeladude jaoks ja andmete kaevandamine.

Arendused 1990-ndatel

- Spetsiaalsed andmelaole orienteeritud andmebaasisüsteemide vajadus ilmnis eelmise sajandi 90-ndate aastate teisel poolel kui andmelaondusega oli tegeletud juba piisavalt kaua selleks, et andmeladude andmemahud oleksid piisavalt suured probleemide tekitamiseks.
- Selgus et olemasolevate andmebaasisüsteemide füüsilised struktuurid ja päringute optimeerimise meetodid ei sobi nii suurte andmemahtude käsitlemiseks ja küll kõigi reeglite järgi “kokku kuhjatud” andmetest ei ole mingit kasu, kuna puudusid vahendid, mis aktsepteeritava ajaga suudaksid analüüsida andmeid ja esitada analüüsi tulemused soovitud kujul.
- Seepärast töötati välja täiesti uue sisemise füüsilise arhitektuuriga süsteemid koos uute päringu optimeerimise meetoditega mis on orienteeritud andmete lugemisele andmebaasist. Lisaks sellele piirati ka oluliselt andmete lisamise ja andmete kustutamise võimalusi. Uuendamise operatsioon puudub nendes süsteemides üldse. Tuntumad andmelao andmebaasisüsteemid, mis tollel ajaperioodil välja töötati olid Sybase IQ ja Oracle DB Data Warehousing.

Arendused 1990-ndatel

- Andmete kaevandamise all mõistetakse sellist tegevust, mis on suunatud suure hulga andmete automaatsele või poolautomaatsele uurimisele ja analüüsile selleks, et leida tähendust omavaid mustreid, mudeleid ja reegleid.
- See andmeanalüüsi suund on otseselt tingitud andmeladude arengust, kuna sellisel määral andmeid, mille alusel andmekaevandamist saab teha on tavaliselt ainult andmeladudes. Selle teooria peamised arendused liideti erinevate äri-intelligentsi toodetega.

Arendused 1990-ndatel

- Suure muudatuse modelleerimisse tõi UML–süsteemide tulek. Suures osas on nad edasiarendused eelmise sajandi 80-ndatel aastatel loodud ERD- ja DFD-mudelite joonistamise CASE-vahenditest. Siiski oli siingi uustulnukaid ja paraku on just nemad olnud innovatiivsemad ja ka edukamad.
- UML-süsteem moodustab tervikliku kogumi erinevaid kirjeldus-elemente, mille abil esitatakse infosüsteemi spetsifikatsioon ja millest hiljem on võimalik genereerida suur osa infosüsteemist.
- UML-süsteemides sisalduvad ka andmete modelleerimise vahendid, mis on küll tihedalt seotud kõikvõimalike muude kirjeldustega (andmevoo-diagrammid, oleku-diagrammid, struktuuri-diagrammid, protsessivoo-diagrammid, kasutuslood jne.).

Arendused 1990-ndatel

- Tuntumad UML-süsteemid on firma Select Inc. Poolt loodud süsteem SELECT ja firma Rational poolt loodud toodete pere.
- Viimase poolt välja töötatud RUP-metoodika (Rational Unified Process) analüüsi läbiviimiseks, infosüsteemi spetsifitseerimiseks ja dokumenteerimiseks kujunes teatavates arendusvaldkondades de facto standardiks.

Areng 21 sajandil

- 21-sajandi algusaastate jooksul ei ole tekkinud juurde uusi suundi andmete modelleerimise ja andmebaasisüsteemide alal.
- Samas on jõudsalt edasi arenenud mitmed juba varem alguse saanud suunad.
- Enamik arenguid on juhitud interneti, kommunikatsioonitehnoloogia, ning nendest tulenevast plahvatuslikult suurenenud kasutajate arvust ja infotöötluste hajumisest ühest keskusest paljudesse erinevatesse punktidesse.

Andmete modelleerimise areng

- Andmete modelleerimise metoodikate ja tehnikate areng ei ole olnud eriti aktiivne ja on väljendunud peamiselt juba varem turul olnud CASE- ja UML-süsteemide uusi omadusi sisaldavate uute versioonide turule tulemisega.
- Täiustatud on analüüsi/spetsifitseerimis-vahendite erinevate mudelite vaheliste seoste kirjeldusi, grupitöö organiseerimise vahendeid (seal hulgas versioonihaldust ja sisu haldust) ja mudelite testimise vahendeid.
- Kogu muutuste suund on dünaamilisuse poole, kus iga kasutajate grupp saab kirjeldada meta-mudelina just endale vajaliku struktuuriga modelleerimiskeskonna ja siis teha sellega tööd maksimaalse efektiivsusega.

Andmete modelleerimise areng

- On siiski üks erand - arenema on hakanud domeen-modelleerimise meetodid ja neid toetavad süsteemid, mis rajanevad mõistete süsteemide loomisele ja domeenide kirjelduste alusel andmemudelite kirjelduste ning programmikoodi genereerimisele.
- Domeeni all mõistetakse mõistete, mõistete vaheliste seoste ja nende andmemudeliteks ning programmikoodiks transleerimise reeglite kogumit.
- Neid reklaamitakse kui “UML-meetodist 10 korda kiirema tulemuse saamise vahendeid” ja on saavutanud selle alal päris häid tulemusi. Domeen-modelleerimissüsteemi abil on suhteliselt lihtne luua endale oma andmete modelleerimise CASE-vahend, mis sisaldab just selliseid komponente ja toetab sellist reeglistikku, mis on sobiv just konkreetsele kasutajale

Andmete modelleerimise areng

- Näiteks kasutab ülemaailmselt tuntud telekommunikatsiooni-seadmeid tootev firma Nokia domeen-modelleerimissüsteemi MetaEDIT+ (MetaCase Consulting) ja arendab ning genereerib selle abil andmestruktuure ja tarkvara toodetavate mobiiltelefonide jaoks.
- Projekt osutus tasuvaks, et vahend on suuteline ära kasutama varasemate arenduste komponente tooteliinide uute toodete tarkvara loomisel.

Arenduse valdkonnad

- Andmete kaitses
- Töökindlus ja käideldavus
- Transaktsioonikäsitus
- Sisu haldus (Content management)
- Andmelaod ja äriintelligents
- Andmete integreerimine
- Andmebaaside haldamine

Andmete kaitse

- Koos interneti ja interneti kaudu kasutatavate rakenduste laialdase levikuga on kiireid ja radikaalseid arendusi nõudnud andmete kaitse.
- Siin võib vaadata kolme arendussuunda – andmetele juurdepääsu reguleerimine, baasisolevate salastamine ja andmete klassifitseerimine turvataseme järgi.
- Andmetele juurdepääsu reguleerimisel on praegu arvestavaim meetod PKI (Public Key Infrastructure) sertifikaadil X.509, versioon 9 (interneti autoriseerimis standard) üle SSL-i (Secure Sockets Layer) põhinev autoriseerimismeetod.

Andmete kaitse

- Andmete salastamisel on juhtivaimaks lahenduseks andmete valikuline krüpteerimine (Selective Data Encryption).
- See tagab tundlike andmete salastatuse ka juhul, kui peaks toimuma andmete autoriseerimata lugemine. Siin rakendatavate printsiipide abil püütakse optimeerida andmete krüpteerimise mahtusid - krüpteeritavateks määratakse ainult tundlikke andmeid. Välja on töötatud krüpteerimisstandardid DES (Data Encryption Standard) ja 3DES (Triple-DES).

Andmete kaitse

- Suuremaks probleemiks andmete turvamisel on olnud andmete horisontaalsete kasutusõiguste kirjeldamine – ligipääsu piiramine vastavalt kasutajate õigustele andmebaasi tabelite erinevatele ridadele.
- Secure Data Sharing - tehnoloogia pakub lahendust selle probleemile. Siin on võimalik kirjed jagada turvatasemete vahel ja vastavalt kasutaja turvatasemele lubada juurdepääsu ühe või teise turvatasemega kirjetele.

Töökindlus ja käideldavus

- Koos pidev-teenusena (24 tundi päevas, 7 päeva nädalas, 365 päeva aastas – 24-7-365) läbi interneti ja ka lokaalvõrgu rakendustena pakutavate virtuaalsete infoteenuste tulekuga on oluliselt kasvanud vajadus andmebaasisüsteemide töökindlusele.
- Infosüsteemi mistahes komponendi (ka andmebaasi) kõrge käideldavuse taseme märkimiseks on võetud kasutusele termin high availability. Selle mõiste alla käivad järgmised arendused.

Töökindlus ja käideldavus

- Andmebaasi klastrid (Database Clusters) on andmebaasi käsitlemise metoodika, kus sama andmebaasiga töötab mitu omavahel ühendatud (interconnected) paralleelserverit (andmebaasimootorit, mis asuvad erinevates füüsilistes arvutites). Klaster-mudel is on võimalik jagada klastris olevate serverite vahel koormust ja mõne serveri töötamast lakkamisel anda tema poolt tehtavad toimingud üle klastrisse ühendatud ülejäänud serverite vahel.
- Samuti on võimalik käivitada klastrisse uusi servereid ilma juba seal töötavaid serverid ümber startimata. Selline metoodika on praeguseks hetkeks rakendatud nt andmebaasisüsteemis Oracle 9i (toote nimi on Real Application Clusters – RAC). Teistel andmebaasisüsteemidel on olemas lihtsamaid klastrite arendusi, kus sama andmebaasiiga ühendatud andmebaasi mootorid töötavad sama arvuti erinevate protsessorite küljes ja automaatika on väiksem.

Töökindlus ja käideldavus

- Pidevalt töötavate (24-7-365) andmebaasisüsteemide kasutamisel on olnud suureks probleemiks ka see, et andmebaasistruktuuride muutmiseks ja muudeks hooldustöödeks tuleb tavaliselt peatada kõigi kasutajate töö andmebaasiga (logoff).
- Uute Online Data Evolution meetodite rakendamisel saab muudatusi teha samal ajal kui toimub aktiivne töö andmebaasiga. See parandab oluliselt juurdepääsu andmetele paljudes missiooni-kriitilistes süsteemides .

Töökindlus ja käideldavus

- Uued meetodid on kasutusele võetud andmete säilimise tagamiseks (Data Protection). Kui seniseid varundus-meetodeid (backup/restore) võib nimetada passiivseteks ja perioodilisteks, siis uusi tehnoloogiaid võib kirjeldada kui aktiivseid ja pidevaid.
- Siin ei toimu andmete varundamine enam range tsüklilisusega vaid on rohkem asünkroonse iseloomuga ja protsessid käivituvad “ise” vastavalt mingite tingimuste täitumisele. Samuti on ka andmete taastamist oluliselt automatiseeritud. Andmete hävimisel on süsteemid ise suutelised analüüsima vea suurust ja taastama hävinud andmed. See jutt siin on siiski palju ilusam kui tegelikkus – sellised süsteemid on alles oma arengutee alguses.

Töökindlus ja käideldavus

- Suur hulk andmebaasisüsteemide tööaja kasulikku ressursi kulub vigaste käsituslausetöötlusele.
- Selle vältimiseks on hakatud ehitama intelligentseid süsteeme, kus süsteemi loojad saavad kirjeldada andmebaasi käsitlemise korralduste veakorrektsiooni reeglid ja seega juba enne käsitluskorralduste andmebaasi mootorile täitmiseks saatmist parsimist teha korraldustele eelkorrektsiooni.
- Need metoodikad on alles oma tee alguses, kuid juba praegu on näha, et nende abil vabastatakse suur hulk seni veatöötluks kulutatud ressursi.

Transaktsioonikäsitlus

- Üsna pikka aega kasutati andmete lukustamisel andmebaasi lehekülje (page locking) lukustamist. Lukustusmeetodi puhul lukustatakse koos kirje lukustamisega terve andmebaasi lehekülg. Põhiliseks miinuseks on see, et koos mingi tuntud kirje lukustamisega lukustatakse teatud hulk tundmatuid kirjeid. Kasutaja jaoks ei ole teada, millised kirjed on ühe andmebaasi lehekülje peal. Kasutaja jaoks on see sisuliselt juhuslik komplekt kirjeid.
- Sellise meetodi kasutamine oli tingitud peamiselt tehnilistest võimalustest. Kui kasutajate hulk või andmebaasikasutamise intensiivsus on väiksed, siis ei ole selles meetodis midagi halba. Kuid koos interneti tulekuga kasvas nii andmebaaside kasutajate arv kui ka kasutamise intensiivsus hüppeliselt.

Transaktsioonikäsitus

- Sellepärast muutus lehekülje lukustusmeetod kohmakaks ja infosüsteemide tööd pärssivaks – “juhuslikult” ja üldsegi mitte vajalikult lukus olevate kirjete suur arv hakkas takistama andmetele juurdepääsu. Kuna vahepeal on arenenud ka tehnilised võimalused, siis on lehekülje lukustamist asunud välja vahetama kirje lukustamine (row level locking). Siin pannakse lukku tõepoolest ainult need kirjed, mille lukustamiseks on põhjust.
- Veel on hakatud kasutama ka andmete nn. “räpase lugemise” (dirty read) metoodikat, kui kasutaja kes on aktsepteerinud sellise lugemise taseme saab lugeda ka lukus kirjeid, kuid ainult kirje seda kujutist, mis viimasel oli enne kirje lukustamist.
- Arendatud on ka kombinatoorseid lukustusmetoodikaid, kus X-, S- ja U-lukkude kombineerimisel “räpase lugemise” metoodikaga saadakse päris häid tulemusi andmete kasutamise konkurentsuse vähendamisel.

Transaktsioonikäsitus

- Vaated (Views) on andmebaasisüsteemide koosseisudes eksisteerinud juba paarkümmend aastat. Andmebaasi jõudluse tõstmise vajadusest tingituna on tehtud muutusi siingi.
- Arendatud on materialiseeritud vaadete (materialized views) kontseptsioon. Kui siiani oli vaade defineeritav kui “virtuaalne tabel” , mis moodustati päringu täitmise ajaks, siis materialiseeritud vaadet hoitakse füüsiliselt andmebaasis ja muudetakse kui muutub mõni vaadet moodustavatest komponentidest.
- See võimaldab hoida oluliselt kokku kallist aega, mis läheb iga kord “kaduma” virtuaalse vaate moodustamisel. Seda metoodikat on mõtet kasutada suure ridade arvuga ja tihti kasutatavate vaadete korral.

Transaktsioonikäsitus

- Nii kaua kui on olnud kasutusel SQL-keel, on olnud andmebaasisüsteemide arendamist läbivaks probleemiks päringute optimeerimine. Cartesia produkti (Cartesian Product) elimineerimine optimeerimisprotsessist on jäänud aegade taha, kuid ka täna leidub palju parandamist SQL-korralduste täituralgoritmide optimeerimise juures.
- Oluliseks on muutunud teooriad, mis võimaldavad suhteliselt kiiresti elimineerida optimeerimisalgoritmide ebaefektiivsed harud. Optimeerimisalgoritmide pidev parendamine on oluline, kuna pahatihti võtab SQL-lausete optimeerimine tunduvalt rohkem aega kui ebaefektiivselt optimeeritud lausete täitmine.

Sisu haldus (Content management)

- Täistekst-indekseerimine ja täistekst-otsing ei ole üldse enam uued leiutised, aga ikka tegeletakse selle arendamisega. Selle on tinginud üha suuremate koguste dokumentide salvestamine andmebaasidesse ja vajadus neid leida dokumendi sisu alusel.
- Tavalisele täistekst otsingule (“fraas”) on lisandunud uued otsingu strateegiad “täistekst: boolean” (full-text boolean), “täpne fraas” (exact phrase), “lähendus” (proximity), “lõigu otsimine” (section searching), väärkirjutised (misspellings), “tõkestamine” (stemming), “mask” (wildcard), “sõnaraamat” (thesaurus), “sõnade ekvivalentsus” (word equivalence) ja scoring.

Sisu haldus (Content management)

- Omaette nähtuseks on andmestruktuurid, mida on hakatud tähistama mõistega “Semi-structured databases” (pool-struktureeritud andmebaasid).
- Tegelikult ei ole siin midagi uut, vaid lihtsalt kahte vana ja hästi tuntud asja on hakatud kasutama uues kombinatsioonis ja on saavutatud uus tulemus. Tihti ei soovita salvestada dokumente vahetult andmebaasi.
- Samas on aga oluline siduda neid dokumente omavahel ja teistegi andmetega.

Sisu haldus (Content management)

- Sellisel juhul lahendatakse asi nii, et dokumentide päised salvestatakse andmebaasi ja nende dokumendipäiste kaudu moodustatakse ka kõik vajalikud seosed teise dokumentide ja muude objektidega.
- Dokumendid ise salvestatakse tavalisse failisüsteemi jagatud teekidesse (Shared Folders). Andmebaasi kirje päisesse kirjutatakse dokumendi URL. Probleemid millega siin peamiselt tegeletakse on seoste püsivus, seoste taastamine ja andmekaitse metoodikad.

Sisu haldus (Content management)

- Loomulikult ei ole mööda saadud ka XML-st ja JSON-st - Lisaks sellele on tekkinud andmebaasid, kus nii andmete kirjeldusi kui andmeid endid hoitakse XML-, JSON- ja sega-kujul.
- Oluliseks on muutumas kõikvõimalikud elektrooniliste maakaartide süsteemid ja sellel baseeruvad infosüsteemid (maakaardid, turismiinfo, maakatastrid, geoloogia süsteemid, ehitiste registrid jne.). Oluline on siin graafilise info kihilisus, graafilise informatsiooni tükeldamine ja graafilise informatsiooni sidumine andmevõrgustikuga.

Sisu haldus (Content management)

- Suuri arvestatavaid tegijaid on palju. Need spetsiaalsed vahendid võimaldavad luua tavalistest geograafilistest infosüsteemidest (GIS) kuni traadita lokaliseerimissüsteemideni ja lokaliseerimist nõudvate E-kommerts teenusteni.
- Seda probleemvaldkonda tähistatakse mõistega location based services.

Andmelaod ja äriintelligents

- Toimub hoogne äriintelligentsi süsteemide arendamine. Täiendatakse meta-mudeleid ja meta-mudelite ehitamise vahendeid. Viimased muutuvad järjest visuaalsemaks.
- Äriintelligentsi süsteemidega liidetakse ka andmekaevandamine ja tekkivad vahendid, mis võimaldavad andmekaevandamise tulemusi pool-automaaitselt ja ka juba osaliselt automaaitselt muuta äriintelligentsi kirjeldavateks meta-mudeliteks.

Andmelaod ja äriintelligents

- Kuna on aru saadud, et nii äri-intelligentsi kasutamise kui ka andmekaevandamise aluseks saab olla võimalikult lai ja korrektselt koostatud andmeladu, siis arenevad kiiresti edasi andmete kollekttsioneerimise ja teisendamise süsteemid.
- Kuna kollekttsioneeritavate andmete voogude hulk kasvab pidevalt ja nende töötlemine andmelao jaoks vajalikule kujule nõuab järjest rohkem ajalist ressurssi, siis arenevad siingi metoodikad ja vahendid andmevoogude ja andmevoogude siseselt paralleeltöötuse suunas.
- Andmete teisendusmudelite kirjelduskeeled muutuvad järjest paindlikumaks ja visuaalsemaks.

Andmete integreerimine

- Hajusandmetöötlaste üks suuremaid probleeme on üksteisest eemalolevate samatähenduslike ja sama struktuuriga või eritähenduslike, kuid protseduurilist seost omavate andmebaaside vahelised andmete ülekanded.
- Esimesel juhul selleks, et hoida kahte või enam andmebaasi omavahel sünkroonis, teisel juhul selleks et liigutada vajalik hulk (koond)andmeid ühest andmebaasist teise.
- Selleks kasutatakse juba pikemat aega andmete replikeerimise meetodeid. Paraku ei ole veel siiani mistahes protsessi toetavat replikeerimissüsteemi loodud.

Andmete integreerimine

- Kui replikatsiooni vahendite arengu algusaastatel murti pead seose terviklikkusest (referential integrity) tingitud replikatsiooniprobleemide üle, siis tänapäeval on peamiseks probleemiks väga võimsaks muutunud transaktsioonide haldamise mehhanismid
- Mille tulemusena ei ole võimalik enam korrata teises serveris kõiki tegevusi just selles samas järjestuses kui seda tehti replikeeritavas serveris.
- Võib olla on siin üldse tegemist valede probleemipüstitustega, kui nii pikkade aastate jooksul (üle 10 aastat) probleemidele korrektset toimivaid lahendusi ei leita. Suurimaks puuduseks on siin replikeerimistegevuste korrektse toimimise järelkontrolli võimaluse puudumine.

Andmete integreerimine

- Viimastel aegadel on paljud kohad, kus seni kasutati mingi keerukusega replikatsiooni, asendatud teatepõhise (MQ - Message Queue) infovahetuse ja sünkroniseerimise mehhanismidega. Siin saadetakse erinevate süsteemide (ka andmebaasimootorite) vahel asünkroonseid teateid just sellisel hetkel kui see vajalikuks osutub.
- Teadetega koos liiguvad andmed ja tegevuskorraldused. Suureks plussiks replikatsioonisüsteemide ees on, et MQ kaudu saadetud teadete järjekord ei muutu, teate adressaadile kohalejõudmine ja töötlemine tagatakse süsteemsete vahenditega. Seega puudub ka järelkontrolli vajadus. Suurt “võimsust” lisab sellele meetodile XML RPC (Remote Procedure Call) kasutamine.

Andmete integreerimine

- Andmete integreerimiseks võib nimetada ka seost oleviku ja mineviku vahel. Väljaarendamisel on tehnoloogiad, mis võimaldaksid mineviku ajahetke etteandmisel näha andmebaasi just sellisena nagu ta tollel hetkel oli – sellisena nagu tollel hetkel olid nii struktuur kui ka andmed.
- Neid metoodikaid ja vahendeid tähistatakse mõistega “Minevikupilt” (Flashback). Sellele tehnoloogiale üritatakse veel liita automaatset SQL-lauset korrigeerimise mehhanismi, mis suudaks kehtiva andmebaasistruktuuri jaoks kirjutatud SQL-korraldused teisendada tolle ajahetke andmebaasi struktuurile vastavaks.

Andmebaaside haldamine

- Kui eelmise sajandi 90-ndatel aastatel olid võtmesõna “ise-haldav” (self-managing) kasutatav vaid mõnede keskmiste ja väikeste andmebaasisüsteemide puhul (tuntuim Centura corp. andmebaasisüsteem SqlBase), siis praeguseks hetkeks on saanud see trendiks ka suurte andmebaasisüsteemide hulgas.
- Lisandunud on veel ka mõiste “ise häälestuv” (self-tuning) “. Ühelt poolt on see aja märk – kõike mida saab automaatseks teha, seda ka automaatseks tehakse. Teiselt poolt on aga suurte andmebaaside haldamine muutunud nii keeruliseks, et andmebaasiadministraatoritelt on vaja osa koormust maha võtta.
- Lisaks sellele on osadel puhkudel arvuti poolt välja kalkuleeritud lahend igal juhul täpsem kui inimese poolt arvatud lahendus.

Andmebaaside haldamine

- Uueks suunaks on andmebaasiserverite ressursside jaotamine ja prioritseerimine vastavalt äriprioriteetidele.
- Peamiselt vaadeldakse siin protsessorite ressursi jagamist erinevate kasutajate, kasutajagruppide ja protsesside vahel, määrates ka paralleelprotsessorite kasutamise tasemed.
- Oluline on protsesside ressursinõudluse automaatne mahasurumine, mis “tuntakse ära” kui “nii kui nii pikalt töötavad protsessid” ja nende arvelt süsteemi üldise käideldavuse tõstmine.
- Jõudsasti arenevad andmebaasisüsteemide distant diagnostiika ja distantshalduse vahendid.