

Duomenų saugyklos

Įvadas

Šiandien daugelis stambių įmonių susidūrė su problema - informacija lyg kažkur yra, jos netgi yra per daug, bet jinai nesuderinta, nesusisteminta, ne visada patikima, jos praktiškai neįmanoma greitai surasti ar gauti. Būtent šitos problemos sprendimui ir skirta duomenų saugyklų (DW - Data Warehousing) koncepcija.

Koncepcijos pagrindą sudaro dvi idėjos:

- duomenų integracija, suderinimas, agregacija,
- išskyrimas transakcinių duomenų ir duomenų skirtų analizei.

DW koncepciją pasiūlė W. H. Inmon 1990 metais. Šiandien tai yra labai populiaru ir sparčiai besivystanti informacinių technologijų sritis. Pagal W. H. Inmoną duomenų saugykla (DW - Data Warehouse) tai yra integruota, dalykiškai orientuota, nekintama, chronologiją palaikanti DB, kuri aprūpina menedžerius ir analitikus patikima informacija, kad jie galėtų priimti efektyvius verslo sprendimus[2].

Yra ir kitų DW apibrėžimų. Pavyzdžiui, R. Kimball apibrėžia DW kaip transakcinių duomenų kopiją, specialiai transformuotą analizei [1].

DW koncepcija nagrinėja pačius duomenis. Po to, kai tradicinė transakcinė (OLTP) sistema realizuota ir pradeda funkcionuoti, jina tampa tokiu pat realaus pasaulio objektu, kaip ir bet kuris kitas gamybinis procesas. Duomenys, kurie yra vienas iš galutinių tokios gamybos produktų, turi tokias pačias charakteristikas ir savybes, kaip ir bet kuris kitas gamybos produktas: galiojimo laiką, sandėliavimo vietą, rinkos vertę ir pan. Būtent iš šitos pusės ir nagrinėja duomenis DW koncepcija. T.y. pagrindinis dėmesys skiriamas ne realaus pasaulio objektų atvaizdavimo ir aprašymo metodams, bet duomenims, kaip savarankiškam dalykinės srities objektui.

Norint teisingai suprasti DW koncepciją, reikia atkreipti dėmesį į tokius momentus:

- DW koncepcija nėra duomenų analizės koncepcija, tai greičiau duomenų paruošimo analizei koncepcija.
- DW koncepcija nenumato tikslinės analitinės sistemos architektūros. Jinai sako apie tai, kokie procesai turi veikti sistemoje, bet ne apie tai, kaip šitie procesai turi veikti.
- DW koncepcija numato vieningo integruoto duomenų šaltinio realizavimą.

DW realizacijos

Galimi tokie DW realizavimo variantai:

- Virtuali DW.
- Duomenų vitrina (Data Mart).
- Globali DW.

Virtuali DW

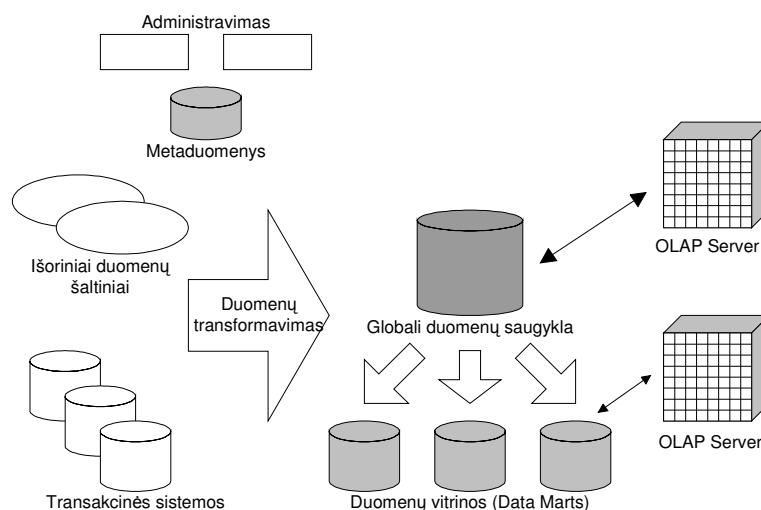
Tokios DW pagrindą sudaro metaduomenys, kurie aprašo informacijos šaltinius, SQL – užklausas ir duomenų atvaizdavimo ir apdorojimo procedūras. Priėjimą prie procedūrų užtikrina tarpinio sluoksnio programinė įranga. Šiuo atveju duomenų pertekliškumas yra nulinis. Galutiniai vartotojai praktiškai tiesiogiai dirba su OLTP sistemomis su visais iš to išplaukiančiais privalumais ir trūkumais. Privalumas yra tame, kad tokiose sistemose visada yra priėjimas prie paskutinių duomenų. Trūkumai: OLTP sistemų našumo sumažėjimas, galimi OLTP sistemų darbo nesklandumai, dėl nekorektiškų analitikų veiksmų.

Duomenų vitrina

Duomenų vitrina – tai yra tam tikru aspektu surištų duomenų bazių rinkinys. Visos į duomenų vitriną įeinančios duomenų bazės turi informaciją apie tam tikrą korporacijos veiklos kryptį. Iš esmės, duomenų vitrina tai yra supaprastintas DW variantas, kuris saugo tik tam tikru požiūriu susijusius duomenis. Tikslinė duomenų bazė maksimaliai priartinta prie galutinio vartotojo ir gali turėti agreguotus duomenis. Duomenų vitrina žymiai mažesnė pagal dydį, negu korporacinė DW. Jai realizuoti nereikia galingos skaičiuojamosios technikos.

Globali DW

Pagrindinė idėja – apjungti DW bei duomenų vitrinų koncepcijas ir naudoti DW kaip vienintelį integruotų duomenų šaltinį visoms duomenų vitrinoms.



DW ypatumai

Nevienalytė programinė aplinka. DW niekada nėra kuriama tuščioje vietoje. Beveik visada galutinis sprendimas bus įvairiarūšis, t.y. jame bus panaudotos autonomiškai sukurtos programinės priemonės. Visų pirma, tai liečia integruoto suderinto duomenų rinkinio formavimą. Duomenys gali būti iš skirtingų tipų duomenų bazių, elektroninių archyvų, elektroninių katalogų, žinytų, statistinių ataskaitų. Projektuojant DW tenka spręsti vieningos, suderintai veikiančios informacinės sistemos kūrimo uždavinį, atsižvelgiant į tai, kad programinės priemonės ir sprendimai gali skirtis. Renkant DW realizavimo priemones tenka atsižvelgti į daugelį faktorių: įvairių programinių priemonių suderinamumas, jų įsisavinimo ir eksploatavimo lengvumas, funkcionavimo efektyvumas ir pan.

Paskirstytas organizacijos charakteris. DW koncepcija numato, kad operacinis analitinis duomenų apdorojimas gali būti vykdomas bet kuriame tinklo taške nepriklausomai nuo to, kur yra pagrindinė saugykla. Nors analitinio duomenų apdorojimo metu duomenys yra tik skaitomi ir nėra sinchronizavimo problemos, vis dėlto efektyvumo dėlei, reikia palaikyti replikavimo mechanizmą skirtinguose tinklo mazguose.

Padidinti duomenų saugumo reikalavimai. Surinkta kartu informacija apie korporacijos vystymo istoriją, jos pasiekimus ir nelaimės, apie bendravimą su tiekėjais ir užsakovais, apie rinkos istoriją ir būklę duoda galimybę analizuoti praeities ir dabartinę veiklą ir taip pat daryti ateities prognozes. Šita informacija yra labai vertinga organizacijai. Jokių būdu negalima leisti, kad jinai patektų į konkurentų rankas. Nes jeigu taip atsitinka, tai visos analitinės organizacijos prognozės bus neteisingos. DW sistemose nepakanka duomenų saugumo, kurį suteikia paprastos DBVS (šitas saugumo lygmuo atitinka oranžinės knygos C2 klasę). Norint užtikrinti reikiamą saugumo lygį, duomenų priejimą reikia kontroliuoti ne tik lentelių ir atributų, bet ir atskirų kortelių lygyje (B1 oranžinės knygos klasė). Taip pat tenka spręsti duomenų saugumo klausimus pakrovimo į DW sistemas metu, duomenų perdavimo tinklu metu [3].

Daugiasluoksnių metaduomenų žinytų būtinybė. OLTP sistemose metaduomenų vaidmuo yra pakankamai ribotas. Metaduomenys tokiose sistemose, kaip taisyklė, laikomi lentelėse – kataloguose. DW sistemose metaduomenų reikšmė yra žymiai didesnė. Tinkamas metaduomenų organizavimas ir pateikimas galutiniam vartotojui yra viena iš pagrindinių sėkmingo DW realizavimo sąlygų. Prieš tai, kai korporacijos menedžeris užduos sistemai savo klausimą, jis turi suprasti, kokia informacija apskritai yra sistemoje, koks yra šios informacijos aktualumas, ar galima šitą informaciją pasitikėti, kiek laiko užtruks atsakymo formavimas ir pan.

Labai didelių informacijos kiekių efektyvaus saugojimo ir apdorojimo poreikis. Jau dabar yra nemažai DW, kurių tūriai viršija 1000 Gb. Pusė korporacijų, kurios naudoja DW sistemas planuoja padidinti jų tūrius iki šimtų Gb. Tokių milžiniškų sistemų problema yra tame, kad išorinės atminties

kaina auga netiesiškai didėjant saugyklos tūriui. Tyrimai TPC-D testais parodė, kad 100 Gb. tūrio duomenų bazei reikia 4.87 kartų daugiau išorinės atminties, negu reikia naudingiems duomenims. Augant DB tūriui, šis koeficientas didėja [3].

Esamos DW rinkos padėties įvertinimas

DW rinką galima suskirstyti į dvi dalis:

- Produktai, skirti DW realizavimui. Pavyzdžiui, Oracle Data Warehouse Builder.
- Konkrečių DW realizavimai, panaudojant pirmosios grupės produktus.

Tarp pirmosios grupės produktų gamybos lyderių galima išskirti JAV korporacijas Oracle ir IBM. Šių korporacijų programinė įrangą yra žinoma ir vartojama visame pasaulyje. Jos dirba duomenų bazių rinkoje jau ne vieną dešimtmetį ir turi didžiulį patirtį bei potencialą. Šiuo momentu jos siūlo savo klientams visą kompleksą programinių priemonių skirtų duomenų analizei. Čia įeina ne tik DW, bet ir OLAP priemonės.

Lietuvoje gaminami tik antros grupės produktai. Pirmosios DW buvo pradėtos gaminti pakankamai neseniai – prieš penkerius metus. Taigi Lietuvos DW rinka yra pakankamai jauna. Tarp Lietuvos įmonių, užsiimančių DW gamyba, galima paminėti UAB “Informacines technologijas” (Oracle atstovas Lietuvoje) ir “Pasparą” (SAS Institute atstovas Lietuvoje) [*pagal 2002 metų duomenis*].

DW poreikis pastaruju metu sparčiai auga. Galima prognozuoti, kad artimiausiu metu atsiras daugiau įmonių, gaminančių DW.

Literatūra

1. Kimball, Ralph. The Data Warehouse Toolkit. John Wiley & Sons, 1996. 416p.
2. W. H. Inmon. Building the Data Warehouse. John Wiley & Sons, 1996. 412p.
3. <http://www.citforum.ru/database/kbd98/glava15.shtml>