

Metadata osaksi tietovaraston sisältöä ja sisällön portinvartijaksi

DW 2.0

Ensimmäisen sukupolven tietovarastot kehittivät hyvin eri lähtökohdista verrattuna nykyisin vallitseviin olosuhteisiin. Niin ymmärrys ja näkemys tietovarastoinnin (Data Warehousing, jatkossa DW) roolista organisaation informaatioarkkitehtuurissa, kuin tekniset reunaehdot ovat muuttuneet ratkaisevasti. Lisäksi tekninen ja menetelmällinen kehitys on johtanut merkittävästi eroaviin tulkintoihin siitä, mitä DW:llä tarkoitetaan ja siitä, miten käyttötarkoitustietokantoihin tukeutuva johtamistiedon tietopalvelu toteutetaan ja ylläpidetään laadukkaasti ja kustannustehokkaasti.

Bill Inmon, tunnustettu tietovarastoinnin asiantuntija ja Enterprise Data Warehousing-opin luoja, on julkaissut määritelmä- ja kuvauskokonaisuuden tiedonhallintamallista jonka on nimennyt DW 2.0:ksi. Tunnusomaista tälle on kokonaisvaltaisuus (hallitsee niin rakenteellisen kuin muotovapaan tiedon), aikajänne (on-line tiedoista – arkistoon) sekä tiedon elinkaaren huomioon ottaminen, metatiedon ja referenssi/viitetiedon rakenteellinen asema mallissa, sekä pyrkimys metatiedon integroimiseen.

DW 2.0 ei sanele matalan tason implementointia, mutta tarjoaa yhteismitallisen viitekehyksen keskustelulle tietovarastoinnista. Inmon on lisäksi hakenut tuotemerkkikirekisteröinnin termille DW 2.0, jolla hän pyrkii selkeyttämään ja yhdenmukaistamaan tietovarastoon liittyvää keskustelua.

DW 2.0:aa voi pitää merkittävänä teknologiarippumattomana referenssimallina informaatioarkkitehtuurikehittämiseksi.

Tässä kirjoituksessa keskitytään metatiedon rooliin tietovaraston dynaamisessa ja staattisessa toteutuksessa.

Metatieto DW 2.0:ssa

DW 2.0:n peruspiirteisiin kuuluu metatiedon jaottelu tekniseen- ja liiketoimintametatietoon. Tekninen metatieto kertoo tiedon tallennuksesta ja mallinnuksesta, ja liiketoimintametatieto tiedon merkityksestä, tulkinnasta ja säännönmukaisuudesta. Molemmilla voidaan ratkaisevasti nostaa tietovaraston semanttista astetta ja automaatiota, sekä parantaa tietovaraston eri sidosryhmien kommunikoinnin perusedellytyksiä.

Tietovaraston liitos perustietojärjestelmien datasisältöön tehdään monesti tähän tehtävään soveltuvilla ETL-työkaluilla (Extract Transform Load). Näiden kehittyminen mahdollistaa aikaisempaa matalammat integraation työmäärät ja kustannukset, voidaan todeta, että tietovarastoin-

nin kustannus/lisäarvo ja riski/hyöty tunnusmerkit ovat ratkaisevasti muuttuneet.

Metatiedon hyödyntäminen ETL integraatiossa

Teknisen- ja liiketoimintametatiedon hyödyntäminen osana dynaamisen ETL-integraatiokerroksen toteuttamista lisää tietovaraston muutosherkkyttä ja parantaa sen elinkaarikustannusennustetta. Lähdejärjestelmistä siirtyvän tiedon tulkinta, validointi ja yhdenmukaistaminen ETL:ssä voidaan perustaa metatiedossa kuvattuihin sääntöihin. Näistä säännöistä on alla olevassa kuvassa käytetty termiä 'requirement'. Kun sääntöjen ylläpitäminen erotetaan niistä toteuttavista ohjelmista, voidaan määrittely- ja muutoshallintaprosessin tulosten kytkeytymistä muutoshallintaan ja itse integraatiototeutukseen merkittävästi tehostaa.

Dynaamisen toteutuksen kääntöpuolet liittyvät muutoshallintaprosessin hallittavuuteen, audit-trailiin ja versionhallintaan. Yksi tapa ratkaista tämä olisi ylläpitää ohjaustieto omassa "staging-production" tauluparissa, joiden väliin voidaan toteuttaa versiotiedon generointi sekä lokitiedon tuottaminen.

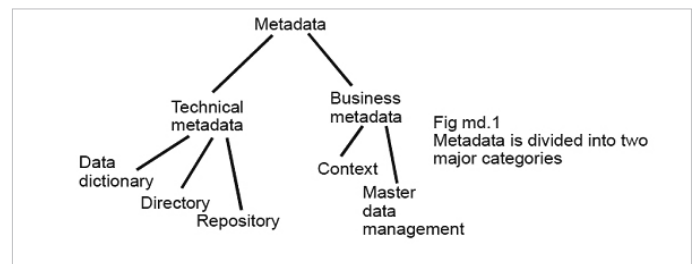
Tässä mallissa esiprosessi ajetaan ennen varsinaista ETL vaihetta, joka tuottaa varsinaisen suoritettavan koodi-instanssin. Kuten yllä todettiin, koko ETL prosessin "configuration" hallittavuuden kannalta täysin ajonaikainen toteutus vaatii hyvin korkeaa hallinta-astetta, mutta eräajona toteutettu versiohallittu esiprosessi yhdistää rakenteen dynaamisuuden edut hyvään dokumentointiin ja hallittavuuteen.

Metatiedon arvo audit trailissä

Koko käsittelyketjun läpinäkyvä seurattavuus on ilmeinen lisäarvo laajasti käsitetystä metatietoarkkitehtuurista. Lähdetiedon lähdejärjestelmä, poimintahetki (sekä mahdollisesti poiminnassa huomioidut ehdot), muunnokset ja esitysmuodon muutokset sekä tietovaraston sisäiset operaatiot ovat kaikki osa sitä metatietoa joka voidaan muodostaa käsittelyvaiheiden sivutuotoksina tai lukea tietovaraston kuvaustietokentistä.



Kirjoittaja toimii konsulttina Ineo Groupin Corporate Performance Management yksikössä.



Kuva lainattu W H Inmonin "ETL in DW 2.0" tekstistä

Välineiden kyky tulkita eri lähderyhmien tuottamaa metatietoa kehittyi edelleen. Kaiken standardisoinnin keskellä, tässä on kuitenkin alue jossa ei vielä ole tunnistettavissa yhteismitallisia muotoja ja rakenteita joilla olisi laajempaa kannatusta. Käyttöjärjestelmien, hakemistopalveluiden, tietokantojen ja eri määrittelytyökalujen metadatan tietomuodot ovat edelleen hyvin moninaiset eivätkä tarjoa tietovaraston laajuiselle tietomäärälle valmista metatiedon hallintamallia- tai rakennetta. Data Vault mallinnusmenetelmän kehittäjä Dan Lindstedt esittää blogissaan ajatuksen metatiedon käsittelemisestä rakenteellisena osana tietovaraston sisältöä: Ottamalla sekä teknisen, että merkitystä kuvaavan (business) metatiedon osaksi tietovarastoa, helpotetaan näiden luettavuutta ja käytettävyyttä, ja lisäksi voidaan varmentaa tiedon käsittely tietovarastoinnissa omaksutuun hyvin käytännöin. Näin myös tekninen metatieto voidaan ottaa avuksi tiedon loppujalosteen oikeellisuutta osoitettaessa.

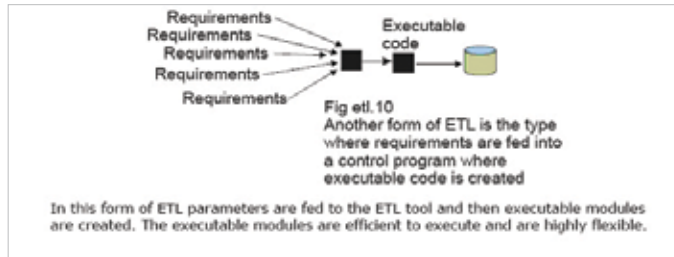
Vastaisuudessa onkin ilmeistä, että kokonaisvaltainen informaatioarkkitehtuuri saa suuremman roolin corporate governance näkökulmasta.

Tällöin myös tietovarastoinnin opit ja tiedonhallinnan hallinta-astetta parantavat piirteet tuottavat uutta lisäarvoa.

Loppujatukset

Tietovarastoinnin menetelmällinen ja tekninen kypsyminen mahdollistavat entistä kustannustehokkaampien toteutusten läpiviemisen, ja tuovat toisaalta tietovaraston hyvät rakenteelliset menetelmät kiinteämmäksi osaksi organisaation liiketoimintatiedon informaationhallintaa. Moni kokee suljettujen tai integroitujen välineiden kautta syntyneen käsitteistön ja datan merkityksen- ja muodon hallinnan haasteelliseksi, ja nämä haasteet kohdataan

aina joko tiedon jalostuskäytön- tai järjestelmäintegraation kohdalla. DW 2.0:ssakin esitetty ajatus kokonaisvaltaisemmasta metatiedon hallinnasta



on harkinnan arvoinen lähestymistapa, jolla tietomistajuutta voi alkaa tukemaan ja toteuttamaan. Toisaalta laajemmalle metatiedon tuottamiselle ja hyödyntämiselle voidaan osoittaa lukemattomia takaisinmaksun muotoja tietovaraston sisäisissäkin prosesseissa ja rakenteissa.



Find out more at:
www.ids-scheer.com/fi

Visionary architecture always requires good building plans!

ARIS Solution for Business-Driven SOA Management helps you to create a transparent service landscape based on business needs.