

Kokemuksia Data Warehouse ratkaisuihin

Antti Kautto, SAS Institute Oy

Menisitkö itse mielelläsi kauppaan jossa kaikki tavarat olisi kasattu pieneen tilaan ja siitä kasasta pitäisi itsepalveluna etsiä haluamiaan tavaroita. Ja lisäksi ahtailla seinustoilla juoksisi ja kulkisi useita samoja tai eri tavaroita etsiviä ihmisiä. Myyjä ei näkyisi missään ja itsestäsi tuntuisi varmasti siltä kuin mitään ei olisi tehty Sinun toivomustesi mukaisesti. Tuskin toista kertaa menisit moiseen puotiin vaikka sen tavarat tuntuisivatkin houkuttelevilta.

Kuitenkin kaikkien ihmisten puheissa kyse olisi tavaratalosta ja joka päivä sinne karrättäisiin uutta tavaraa entisten joukkoon, samaan kasaan.

Jos viereisellä tontilla olisi siisti puhdaslinjainen rakennus josta löytäisit juuri kaikki ne tavarat joita itse tarvitsisit, kumman valitsisit ostospaikkaksesi. Todennäköisesti useimmat meistä valitsisimme tämän jälkim-

mäisen. Ja samassa kaupassa tuntuisi olevan muillekin juuri heidän tarvitsemiaan tavaroita. Eikä mitään turhaa ja tarpeetonta, ei kaikkia taivaan ja maan väliltä vaan oikeita ja vielä sopivasti ryhmiteltynä niin että saa kassan täyteen, jopa molemmilta puolista käytävää yhtä aikaa.

Silti toiset juoksevat, varmaan heillä on vielä kiireempi mutta itsekin toteat ulos tullessasi ettei aikaa mennyt kuin hetki, sillä kaikki ne mitä etsit, löysit heti ensi hyllystä ovesta tullessasi. Ja ulko-ovia oli joka puolella kuten sisäänkäyntejäkin. Toiset kävijät ottivat suoraan hyllystä ison valmiin laatikon ja kun parkkipaikalla näit vieruskaverisi avaavan tällaista, paljastuikin sen sisältävän laajan valikoiman kaikkia juuri samoja tavaroita joita itsekin olit hyllyistä etsiskellyt. Tällaisessa laatikossa ne kuitenkin olivat valmiiksi tarjoiltuna, juuri käyttäjän kannalta ajateltuna. Tällaisen

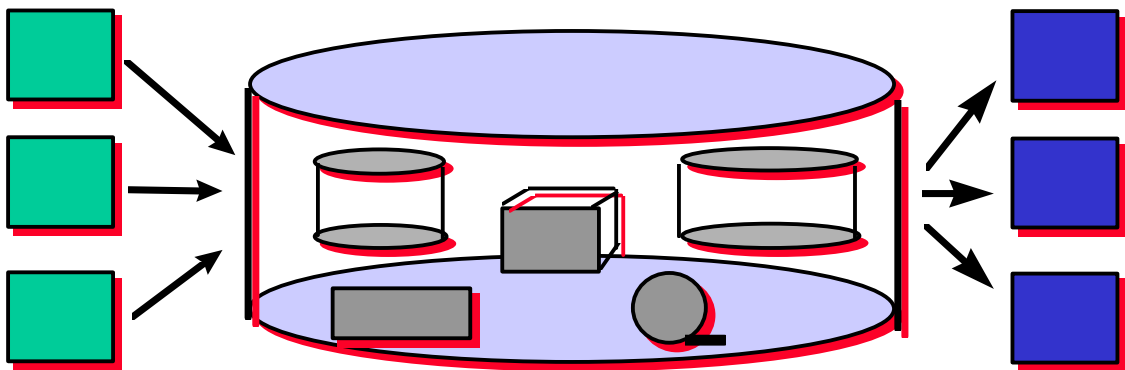
laatikon otankin ensi kerralla. Lisäksi tavarat eivät maksaneet yhtään enempää ja ennenkaikkea kokonaisaikaa säästy.

Kun seuraavana aamuna kävit hakemassa vielä pari pikku tavaraa, huomasit että kaikki oli tuoretta ja siististi tarjolla. Varasto oli täytetty yöllä ja paikat järjestelty uuden päivän tarpeisiin.

Kumman itse valitsisit, tietojen varaston josta oikean tiedon löytäminen tuntuu olevan enemmän kyselijän vastuulla vaiko oikean Data Warehousen josta tuorein tieto on jo heti ovelta löydettävissä ja valmiina hyödynnettäväksi.

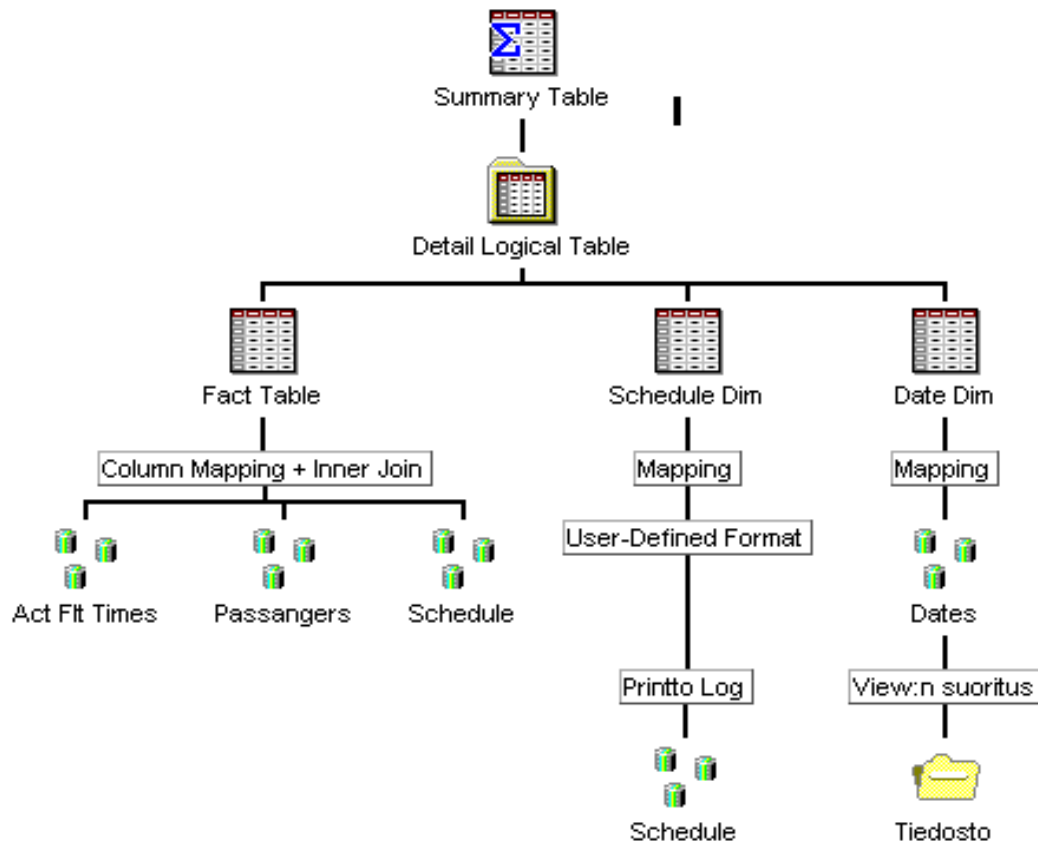
Olen itse eri tietovarastohankkeissa aika usein törmänyt juuri tällaiseen vaihtoehtoasetteluun. Jos olisi kyse edelläkuvatusista kaupan valinnasta olisi aika todennäköistä että jokainen valitsisi selkeän ja tarkoitusta varten rakennetun tiedon tavaratalon. Kuitenkin aika monessa

DATAWAREHOUSE ei ole pelkkä tietokanta



Sinne tulee tietoa useista eri tietolähteistä ja tämä tuleva tieto yhdistetään ja muokataan käsittelyprosessissa vastaamaan tiedon tarvisijoiden eli **liiketoiminnan** kyselytarpeita.

=> PROSESSI ei TIETOKANTA



tapauksessa mietitään ratkaisua puhtaan tekniikan kannalta. Otetaan vaan kaikki tieto, laitetaan se yhteen isoon laatikkoon ja päälle leima "Data Warehouse" ja jaellaan käyttäjille kyselytyökaluja kehoituksella "antaa palaa".

Tästä on hyviä kokemuksia kun käyttäjänä on ollutkin todellinen asiantuntija, tiennyt tiedon ja käsitteiden tarkoitukset ja syntymekanismit. Tällainen käyttäjä on ollut tuloksiinsa tyytyväinen. Kaikki työkaveritkin pitivät häntä arvossa. Tällainen asiantuntija osaa rakentaa työkalulla itselleen ja ehkä muillekin hyviä kyselyjä. Niillä saa tietoa ja niiden pohjalta on hyvä tehdä päätöksiä ja johtaa yritystä.

Usein tällaisissa tapauksissa on tiedon laadussa ongelmia. Parhaatkaan lähdejärjestelmät eivät aina tuota täyttä laatua vaan välillä syntyy suttua ja sekunda. Ei laaduttomuutta kukaan tee, sitä vaan syntyy ja se kulkeutuu kuin tarttuva tauti, päätöksiin asti. Torjuakseen virheellisyyksiä, tällainen asiantuntija rakentaa tarkistus-rutiineja. Näillä pystytään paljastamaan puutteita tiedon laadussa ja kor-

jaamaan sitä kautta lähdejärjestelmissä olevia aukkoja.

Ja seinän takana toinen samanlainen asiantuntija rakentaa omia rutiinejaan, samoja ja erilaisia. Ne tehdään ennenkaikkea omiin tarkoituksiin ja käsittelyyn otettavien lähtötietojen mukaanotto ja kaikenlainen käsittely saatta hyvinkin paljon poiketa seinän toisella puolella samoja puuhia tekevän kollegan vastaavista logiikoista. Kuukausipalaverissa sitten keskustellaan pitkän siitä, kummalla on oikeat luvut ja kumman taikka kenenkä lukuihin pitäisi voida luottaa ja päätöksiä sen mukaan tehdä. Ja valmistuksessa päätökset tehdäänkin ihan eri lähtökohtien perusteella kuin myynnissä. Molemmilla on tuki lähtökohtana kannattavuus mutta tarkoituksiperät kannattavuuden laskennassa lomituvat vain osin toistensa kanssa.

Miten asiat olivat ennen. Silloin patruunalla oli lehti jo aamukahvilla pöydässä tarjolla ja konttorille tullessaan toimistopäällikkö kertoi heti päivän tärkeimmät liiketoimintatiedot ja muut merkittävät asiat. Patruuna siitä kysymään, miksi siellä asiat ovat

näin ja miksi tuolla niin. Ja jos toimistossa asioita ei heti osattu vastata tai ei tiedetty, tiedot etsittiin porautumalla tuota pikaa. Ja jo ennen lounasta oli monta tärkeää asiaa ehditty päättää ja saattaa toimenpiteisiin. Aivan varmasti jotkut tekivät päätöksiä "mutu" menetelmällä mutta aika usein ihan oikean analyttisen päätöksentekotiedon pohjalta. Sen aikainen Data Warehouse vain toimi niin, miksei sama idea voisi edelleenkin olla, ihan samalla yleisperiaatteella, toimivan Data Warehousen lähtöajatuksena. Siis tieto tarjolla käsiteltynä tarpeen mukaan.

Tällaisia Data Warehouseja on tehty. Ja niistä on saatu hyviä kokemuksia. Tietomassat kasvavat hallitusti, tiedolla on oikea paikkansa ja ja on jäsennelly tarpeen mukaan. Lajiteltu, yhdistetty, putsattu ja muokattu. On valmista summaa, lajiteltua hyllyä oikealle ja vasemmalle, ylöspäin, alaspäin ja kolmanteen asteeseen asti. Joillekin tieto tarjotaan laatikossa, toisille jopa väreillä varustettuna ja kotiin kuljetettuna.

Virheä valo kadulla sallii turvallisen etenemisen. Työpaikalla

LIKETOIMINTAPROSESSI

- oper
- as.tunt.
- johto

Markkinointi

- oper
- as.tunt.
- johto

Myynti

- oper
- as.tunt.
- johto

Tuotanto

- oper
- as.tunt.
- johto

Taloushallinto

- oper
- as.tunt.
- johto

Tuotekehitys

Toimintokohtaiset perusjärjestelmät

vihreä valo osaston mittarissa kertoo heti aamulla työaseman avautuessa että siellä asiat ovat hyvin. Asioita ei ole kerrottu hyviksi järjestelmään vaan ne ovat yhteisesti sovittujen liiketoimintaperiaatteiden mukaisesti analysoituna oikeiksi ja hyviksi todettuja. Ei epäilystäkään etteikö siellä asiat ole kunnossa, yhtiön tavoitteiden mukaisesti ja luotettavasti. Ja jos asiakkaan kannattavuuden valo ruudulla on punainen, sen takana on oikeaa ongelmaa ja huolta. On porauduttava juuri sinne ja katsottava mistä nyt kiikastaa. Siitä analyysitiedot palaveriin ja päättämään mitä tehdään tilanteen muuttamiseksi.

Tällainen Data Warehouse suunnitellaan ja pitääkin suunnitella tiedon tarvisijoille, palvelemaan liiketoimintaa ja sen johtamista. Koko yrityksen kannalta, osaston kannalta ja jokaisen tiedon tarvisijan kannalta. Näin on tehty monessa hankkeessa, tukemaan liiketoiminnan tuloksen tekoa. Ja tällä tavalla tehdyissä DW-hankkeissa on onnistuttu hyvin kun kokemus onnistua on käytetty hyväksi.

Onnistumisen edellytyksenä on kuitenkin aina ollut yrityksen johdon

selkeä tuki ja tavoiteasettelu. Mittareista on valittu ensin tärkeimmät, toteutettu ne siivuina, saatu toimimaan joka päivä, 24 tuntia ja luotettavasti. Ensimmäisistä tulokista on jo voitu olla ylpeitä ja tietoa on heti käytetty hyväksi. Seuraavan siivun toteutumista jo odotetaan innokkaana sillä elämä on helpottunut Data Warehousing ajattelun myötä. Asioita katsotaankin nyt liiketoiminnan näkökulmasta eikä kukin omasta pilttuustaan.

Data Warehousen tiedon lataukseen on kirjoitettu yhtiön liiketoiminnan sääntöjä. Ja samalla Data Warehousesta on tehty palvelin. Se ei enää varastoikaan tietoa vaan se pystyykin jalostamaan tietoa, sääntöjen edellyttämällä tavalla. Ja tarjoamaan tiedon lisäksi myös tietoa sisällöstään eli tietoa tiedosta, metadataa.

Kuten entisajan patruunallekin, tietoa tarjottiin, kaikkea ei tarvinnut kysyä. Mutta kenellekään ei olisi edes tullut mieleen pitää johtajalle tunnin tai kahden esitelmää kaikesta mahdollisesta tiedosta, tai antaa 10 sentin raporttia tutkittavakseen. Miksi nykyaikaisen Data Warehousen käyttäjän pitäisi tyytyä yhtään vähempään. Eli tiedon tarvisija voi pidät-

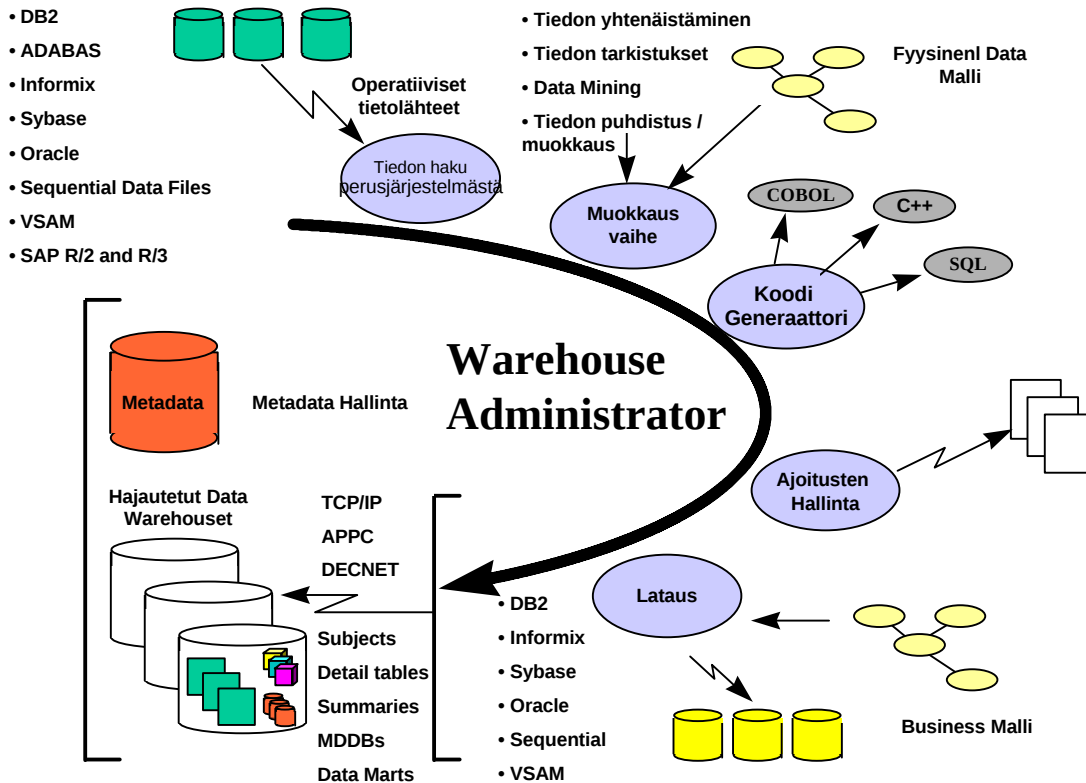
täytyä vain oleellisen eli liiketoimintaan ja sen edistämiseen tiedon etsimisen sijaan.

Tiedon laadun kannalta oikein suunniteltu ja rakennettu Data Warehouse on oivallinen keino selvittää myös poikittaisia riippuvuuksia ja analysoida tietoa pidemmän ajan kannalta katsottuna. Kun kaikelle tiedolle, sisään tulevalle ja DW:ssä oleville tietoelementeille on kirjattu tiedon omistaja, kaikki DW:n prosessit voivat suoraan tiedottaa puutteista ja poikkeamista suoraan asianomistajille, vaikka sähköpostilla. Jälleen vapautuu aikaa oleelliseen, kun ei turhaan tarvitse tutkia, onko kaikki tietovirta mennyt ja toteutunut asianmukaisesti. Jos ei ole huomauttamista, ei yksityiskohtiakaan kannata tutkia.

Toiset käyttävät internettiä käyttöliittymänään, toiset haluavat enemmän toiminnallisuutta ja valmiita työkaluja analysointiinsa. Data Warehouse ei aseta eikä saakaan mitenkään rajoittaa hyödyntämisen välineitän tai tapoja, avoimuutta ja riippumattomuutta on oltava tarjolla.

Miten tällainen hyvä Data Warehouse sitten hallitaan kaikki 24 tuntia vuorokaudessa, jok'ikinen päivä?

Data Warehousen hallinta ja organiointi



Aiemmin Data Warehousen koko kokonaisuuden hallinta on ollut tapahtumakäsittelysovelluksia hankalampaa. Tietovirtojen käsittelyyn kun ei työvälineitä ole ollut tarjolla. Prosessi kuljettaa tietoa on yhtä mutkikas kuin prosessi kuljettaa kaikki paperikoneen osat eri tuotantolaitoksista ja alihankkijoilta kokoonpanojen kautta valmiiksi toimivaksi laitokseksi ostaja-asiakkaan tehdaspaikkakunnalle. Sen hallinnan toteuttamiseen on käytetty projekteissa tuhansia työtunteja, metreittäin dokumentaatioita ja projektien hallintaan kehitettyjä ja tarkoitettuja työvälineitä. Pitkän tapahtumaketjun ja tietovirran hyvä hallinta on Data Warehousen toteuttamisen oleellinen avainasia. Ja hallinnan osaa parhaiten sellainen jolla on kokemusta, välineitä ja kumppaneita.

SAS Institute on työskennellyt jo yli 20 vuotta tiedon jalostamiseksi. Tilastomatemaattinen tausta auttaa ymmärtämään tiedon riippuvuutta, säännöllisyyttä ja epäsäännöllisyyden taustaa. Sen pohjalta ovat syntyneet tämän päivän Data Warehousen oleelliset komponentit ja tuotteet. Näillä on tehty tuhansia onnistuneita ratkaisuja, mutta ei yksin SAS Instituten toimesta vaan yhdessä sertifioitujen kumppanien ja asiakkaitten omien organisaatioiden kanssa.

Nyt uusimpina ovat kehitetyt Data Mining Enterprise Miner ja Data Warehouse Administrator. Näillä varmistetaan se että asiakaskannattavuutemme analysointi pohjaa oikeaan ja todelliseen asiakkaiden käyttäytymiseen, siis ostokäyttäytymisen ja kulukohdistusten lopputulosten datan arvojen analysointiin eikä mutu-

pohjalta ennakkoon rakennettuihin bitmappeihin ja muihin indeksipohjaisiin luokitteluihin. Ja että tiedon keruu toimii ja DW:n käsitteet ovat metadatan kautta kaikkien käytettävissä. Näillä hyvät ja toimivat Data Warehousest pystyttäneet yritykset ovat saavuttaneet sen merkittävän kilpailuedun kansakilpailijoihinsa nähden joilla yhä kiristyvän kilpailun haasteet on pystytty kääntämään voitoiksi.

Lisätietoa aiheesta: Antti Kautto
(antti.kautto@ssf.sas.com), SAS Institute Oy