

Liiketoimintatiedon hallinta
Kandidaatintyö

TIIVISTELMÄ

Lappeenrannan teknillinen yliopisto

Tietotekniikan osasto

Antti Hamara

Kandidaatintyö

Liiketoimintatiedon hallinta

2008

31 sivua, 3 kuvaa, 1 taulukko

Tarkastaja: Professori, FT Juha Puustjärvi

Hakusanat: Liiketoimintatiedon hallinta, tietovarasto, OLAP, tiedonlouhinta

Nyky-yritysten liiketoiminta on muutamassa vuodessa muuttunut todella tietointensiiviseksi. Erityisesti ilmiö on nähtävissä korkean teknologian aloilla, mutta tiedon merkitys lisääntyy myös perinteisillä aloilla. Tieto ja sen oikeanlainen hyödyntäminen on noussut yhä tärkeämmäksi kilpailutekijäksi, mutta samaan aikaan valtavalla tahdilla kasvava tietomäärä vaikeuttaa hyödyllisen tiedon löytämistä. Ilman teknisiä apuvälineitä tietotyöläiset ja johtajat hukkuvat tiedonpaljouteen ja samaan aikaan entistä suurempi osa heidän työajastaan kuluu tärkeän tiedon etsimiseen.

Liiketoimintatiedon hallinta tarkoittaa yrityksen omista prosesseista saatavan sisäisen tiedon ja yrityksen ulkopuolelta, kuten kilpailijoilta tai asiakkailta, saatavan ulkoisen tiedon keräämistä, analysointia ja jalostamista yrityksen liiketoiminnan tehostamiseksi. Liiketoimintatiedon hallinnan tavoitteena on etsiä valtavasta tietomäärästä yrityksen liiketoiminnan kannalta tärkeä tieto ja tarjota se yrityksen eri tasoilla toimivien työntekijöiden ja päättäjien työskentelyn tueksi.

Tämän työn tavoite on antaa lukijalle käsitys mitä liiketoimintatiedon hallinnalla tarkoitetaan ja mitä sen avulla voidaan tehdä sekä mistä lähteistä tietoa saadaan. Lisäksi työssä selvitetään liiketoimintatiedon hallinnan historiaa ja kehittymistä. Työn pääpaino on liiketoimintatiedon hallintaan liittyvissä teknologioissa. Lukijalle pyritään antamaan käsitys liiketoimintatiedon hallintaan liittyvistä tärkeimmistä teknologioista sekä niiden merkityksistä koko liiketoimintatiedon hallintaprosessissa.

ABSTRACT

Lappeenranta University of Technology
Department of Information Technology

Antti Hamara

Thesis for the Degree of Bachelor of Science in Technology

Business Intelligence

2008

31 pages, 3 figures, 1 table

Examiner: Professor, Ph.D. Juha Puustjärvi

Keywords: Business Intelligence, Data Warehouse, OLAP, Data Mining

Knowledge intensity of modern businesses has grown hugely during past few years. This phenomenon is most clearly visible in high technology business but knowledge has grown its magnitude in more traditional businesses also. Knowledge and its right usage is ever more meaningful competitive factor but in the same time the huge amounts of new data available makes it harder and harder to find relevant information. Without right tools knowledge workers and managers drown in the data stream and at the same time ever more of their working time is spent on exploring the data masses searching for the relevant information.

Business Intelligence is collecting, analyzing and taking advantage of data and using gathered information to improve business performance. This data is available from internal information systems and external sources such as the Internet, customers and competitors. Objective for Business Intelligence is to find relevant information from huge amounts of data from several sources and to provide this information to managers to support their decision making.

This thesis presents the basics of Business Intelligence. It is objected to explain what Business Intelligence is, for what it can be used for and which data sources it can utilize. Origins and development of Business Intelligence are also discussed. Main focus is to present the key technologies behind Business Intelligence and their meanings for Business Intelligence process.

SISÄLLYSLUETTELO

1	JOHDANTO	2
1.1	Työn tausta	2
1.2	Työn tavoitteet ja rajaukset	3
1.3	Työn rakenne	4
2	LIIKETOIMINTATIEDON HALLINTA	5
2.1	Johdatus liiketoimintatiedon hallintaan	5
2.2	Liiketoimintatiedon hallinnan historia ja kehitys	7
2.3	Liiketoimintatiedon lähteet ja muodot	8
3	LIIKETOIMINTATIEDON HALLINNAN TEKNOLOGIAT	10
3.1	Liiketoimintatiedonhallinnan arkkitehtuuri	10
3.2	Tietovarasto ja tietovarastointi	12
3.3	OLAP (On-Line Analytical Processing)	15
3.3.1	Coddin määritelmät ja säännöt	16
3.3.2	FASMI-testi	19
3.3.3	OLAP-operaatiot	20
3.4	Tiedonloushintaa	22
4	LIIKETOIMINTATIEDON HALLINNAN SOVELLUKSET	23
4.1	Liiketoimintatiedon hallinnan sovellusten osa-alueet	23
4.2	Liiketoimintatiedon hallinta ja muut tietojärjestelmät	23
5	JOHTOPÄÄTÖKSET JA YHTEENVETO	25
	LÄHTEET	27

1 JOHDANTO

1.1 Työn tausta

Yhä useamman yrityksen liiketoimintaan liittyy keskeisenä osana liiketoimintatiedon hallittu hyödyntäminen. Liiketoimintatieto voi olla yrityksen sisäistä tietoa tai ulkoista tietoa. Yrityksen sisäistä tietoa saadaan yleensä yrityksen omista tietojärjestelmistä kuten ERP- eli toiminnanohjausjärjestelmät (Enterprise Resource Planning) ja CRM- eli asiakkuuksienhallintajärjestelmät (Customer Relationship Management). Sisäinen tieto liittyy yrityksen liiketoimintaprosesseihin ja on siis esimerkiksi tietoa maksutapahtumista. Ulkoisen tiedon lähteiksi taas voidaan lukea lähes kaikki yritykseen, sen toimialaan tai liiketoimintaan liittyvät asiat kuten kilpailijat ja asiakkaat. Ulkoista tietoa ovat esimerkiksi asiakkaiden yhteydenotot yritykseen sekä kilpailevien yritysten www-sivuilla oleva tieto. Aikaisemmin yritysten ongelmana oli oleellisen ja liiketoiminnan kannalta hyödyllisen tiedon vähäinen saatavuus. Nykyään yritysten sisäisten tietojärjestelmien tuottaman ja yritysten saatavilla olevan ulkoisen tiedon määrä on valtava, joten tärkeimmäksi ongelmaksi onkin muodostunut yrityksen liiketoiminnan kannalta oikeasti tärkeän tiedon löytäminen kaiken tietomäärän seasta ja tämän tiedon tehokas jalostaminen ja hyödyntäminen.

Liiketoimintatiedon hallinta (Business Intelligence) on työkalu tärkeän tiedon löytämiseen, analysointiin ja hyödyntämiseen. Liiketoimintatiedon hallintaan tarkoitettujen ohjelmistojen markkinat ovat kasvaneet jo useita vuosia eikä kasvun hidastumista ole odotettavissa lähivuosina. Mielenkiintoisten tulevaisuudennäkymien ja henkilökohtaisen kiinnostuksen takia päätin paneutua kandidaatintyössäni liiketoimintatiedon hallintaan.

1.2 Työn tavoitteet ja rajaukset

Tämä työ on selvitystyö, jonka tavoitteena on selventää termin liiketoimintatiedon hallinta merkitystä sekä siihen liittyviä teknologioita. Englanninkieliselle alkuperäisstermille ”Business Intelligence” vahvin suomenkielinen käännös on ”liiketoimintatiedon hallinta”. Kuitenkin englanninkielisen termin tavoin suomenkielisenkään käännöksen merkitykselle ei ole olemassa yhtenevää linjausta. Termin merkitys voidaan näkökannasta riippuen käsittää kahdella tavalla. Ensimmäinen kanta painottaa liiketoimintatietoa. Tämä kanta on selkeästi esillä sekä ulkomaisissa että suomalaisissa johtamista ja kauppatieteitä käsittelevissä kirjoituksissa. Liiketoimintatietoa ja sen hallintaa käsitellään johtamisen välineenä. Toinen kanta korostaa tiedonhallintaa. Tätä kantaa edustavat tietotekniset julkaisut ja liiketoimintatiedon hallinta nähdään nimen omaan tietojenkäsittelynä erilaisten tietojärjestelmien avulla. Näiden kahden näkökannan raja voidaankin helposti vetää käyttäjän ja tietojärjestelmän väliseen rajapintaan. Tietoteknisessä näkökannassa keskitytään teknologioihin ja työkaluihin, joilla suuresta määrästä dataa voidaan jalostaa hyödyllistä tietoa. Kaupallinen näkökanta käsittelee tämän tietojärjestelmästä ulos saatavan tiedon hyödyntämistä liiketoiminnassa.

Työssä annetaan johdatus liiketoimintatiedon hallintaan ja tarkastellaan liiketoimintatiedon hallinnan historiaa sekä kehittymistä nykyiseen muotoonsa. Suuri osa työstä käy läpi liiketoimintatiedon hallinnan taustalla olevia teknologioita ja niiden merkitystä koko liiketoimintatiedon hallinnan prosessille. Työn aihepiiri kattaa liiketoimintatiedon hallinnan käsitteestä tärkeimpiin teknologioihin saakka. Työssä tutkitaan mitä lähteitä ja muotoja liiketoimintatiedolle on ja perehdytään toimintoihin, joita tiedon jalostamiseen voidaan käyttää. Tarkastelu rajataan liiketoimintatiedon hallinnan teknologiaan. Työssä ei siis selvitetä kuinka järjestelmän tuottamaa tietoa voidaan hyödyntää yrityksen liiketoiminnassa.

Näin voidaan muodostaa työlle seuraavat tutkimuskysymykset:

- 1. Mitä on liiketoimintatiedon hallinta?*
- 2. Mitkä ovat liiketoimintatiedon lähteet ja muodot?*
- 3. Mitkä ovat liiketoimintatiedon hallinnassa käytettävät teknologiat?*

1.3 Työn rakenne

Työ on jaettu kolmeen osioon. Ensimmäisessä osiossa eli luvussa kaksi annan johdatuksen liiketoimintatiedon hallintaan. Luvussa käsitellään liiketoimintatiedon hallintaa käsitteenä, sen tarkoitusta ja tavoitteita sekä lähteitä, joista tietoa saadaan. Ensimmäisessä osiossa tarkastellaan myös liiketoimintatiedon hallinnan historiaa, sen kehittymistä sekä sen nykyistä asemaa muiden tietojärjestelmien joukossa.

Toisessa osiossa tarkastellaan liiketoimintatiedon hallinnan taustalla olevia teknologioita ja niiden liittymistä toisiinsa. Liiketoimintatiedon hallinta koostuu kolmesta pääteknologiasta, joista kustakin esitellään perustiedot ja niiden merkitys liiketoimintatiedon hallinnassa.

Kolmannessa osiossa käydään lyhyesti läpi liiketoimintatiedon hallinnan kytkökset muihin tietojärjestelmiin. Tässä osiossa tarkastellaan myös kuinka liiketoimintatiedon hallinta tulee integroitumaan yritysten tietojärjestelmiin tulevaisuudessa.

Viimeisessä osiossa esitellään johtopäätökset ja yhteenveto. Tässä osiossa kokoan liiketoimintatiedon hallinnan tärkeimmät asiat ja teknologiat selkeäksi kokonaisuudeksi ja selvennän hieman mihin suuntaan kehitys liiketoimintatiedon hallinnassa on menossa.

2 LIKETOIMINTATIEDON HALLINTA

”Tieto on valtaa” sanoi englantilainen filosofi sir Francis Bacon (1561-1626). Tämä viisaus on tullut entistä ajankohtaisemmaksi nykypäivän yritystoiminnassa, mutta yhä useampi yritys kohtaa ongelmia vanhan viisauden hyödyntämisessä. Kalakotan ja Robinsonin (2001, 349) mukaan haaste on, kuinka muuttaa valtava määrä arvokasta tietoa, joka sijaitsee yrityksen ohjelmistojen, tiedostopalvelimien ja tietokantojen syövereissä, uusiksi ansaintamahdollisuuksiksi.

Yksi keino tähän on kerätä liiketoiminnan kannalta oleellinen tieto keskitettyyn paikkaan, jossa tietoa voi tarkastella ja analysoida erilaisin työkaluin. Muodostetut analyysit saatetaan tämän jälkeen päätöksentekijöiden haltuun, jotta heillä olisi mahdollisimman paljon liiketoiminnan kannalta oleellista tietoa hallussaan tehdessään liiketoimintaa koskevia päätöksiä. Liiketoimintatiedon hallinta onkin datan ja tiedon jalostamista päätöksenteon tueksi.

2.1 Johdatus liiketoimintatiedon hallintaan

Liiketoimintatiedon hallinta (Business Intelligence) on suhteellisen uusi tutkimusalue ja tästä johtuen IT-alan tutkimuksessa ei ole päästy yhteisymmärrykseen termin määritelmästä ja laajuudesta. Liiketoimintatiedon hallintaan liittyvät teknologiat ja prosessit ovat myös olleet suuren mielenkiinnon kohteena. Liiketoimintatiedon hallinnan kehitystä on pääasiassa vieneet eteenpäin markkinat ja ohjelmistotalot ja monien muiden suurta huomiota saavuttaneiden ilmiöiden (kirjoittajien alkuperäinen ilmaus – Buzz Word) tapaan liiketoimintatiedon hallinnan varsinainen merkitys on jäänyt epäselväksi. Suurin osa liiketoimintatiedon hallintaan liittyvästä kirjallisuudesta on peräisin IT-teollisuudesta ja akateemista tutkimusta aiheesta on vähän. Tämän takia kirjoittajat peräivät keskustelua liiketoimintatiedon hallinnan merkityksestä ja laajuudesta, sen peruskonseptin merkityksestä ja pääkomponenteista. (Jagielska et al., 2006, 228)

Liiketoimintatiedon hallinta voidaan käsittää prosessina, joka muuttaa datan informaatioksi ja edelleen tiedoksi (Golfarelli et al., 2004, 1). Negashin & Greyn (2003, 3190) mukaan liiketoimintatiedon hallinta yhdistää datan keräämisen, datan varastoinnin ja tietojohdamisen (Knowledge Management) analyttisiin työkaluihin tarkoituksena esittää monimutkainen ja liiketoiminnallisesti tärkeä tieto suunnittelijoille ja päätöksentekijöille. Hannula (2005) taas kuvaa liiketoimintatiedon hallinnan seuraavasti - ”Liiketoimintatiedon hallinnalla tarkoitetaan toimintaa, jossa systemaattisesti kerätään liiketoimintaan ja liiketoimintaympäristöön liittyvää tietoa, tulkitaan ja analysoidaan sitä, arvioidaan tiedon merkitystä ja käytetään analysoitua tietoa päätöksenteon tukena”

Kuten monet muut yleiset konseptit, myöskään liiketoimintatiedon hallinnalle terminä ei ole selkeästi määriteltyä merkitystä. Termin merkitys muuttuu myös näkökulman vaihtuessa, joten joissain yhteyksissä liiketoimintatiedon hallinta käsitetään datan raportointina ja visualisointina. Tietokantatoimittajat korostavat datan muokkaamista ja integrointia. Analysointityökalujen toimittajat taas korostavat tilastollista analyysiä ja tiedonlouhintaa (Data Mining). Erilaisista näkökulmista johtuen liiketoimintatiedon hallinnalla on monta puolta. Yleisesti liiketoimintatiedon hallinta voidaankin kuvailla yrityksen arvokkaan varallisuuden, raakadatan, keräämisenä, ymmärtämisenä, analysointina ja muokkaamisena sellaiseen muotoon, jota voidaan hyödyntää yrityksen liiketoiminnan tehostamiseen. (Azvine et al., 2005, 215)

Tässä työssä liiketoimintatiedon hallintaa käsitellään kuitenkin työkaluina eli teknologioina, tietojärjestelminä ja sovelluksina, joilla liiketoimintatietoa hallitaan ja käsitellään ja joiden avulla merkityksellinen tieto saadaan sitä tarvitsevien ulottuville. Abukari & Jog (2002, 45) määrittelevät nämä työkalut seuraavasti: Liiketoimintatiedon hallinnan työkalut auttavat organisoimaan ja paketoimaan tietoa niin että johtajat pystyvät dynaamisesti tutkimaan, arvioimaan ja analysoimaan liiketoimintaa koskevaa tietoa useista eri lähteistä.

Jagielska et al. (2006) määrittelevät liiketoimintatiedon hallintajärjestelmän yhdistelmäksi dataa, informaatiota, prosesseja, työkaluja ja teknologioita, jotka tarjoavat päätöksentekijöille näkemystä tai tietoa, ja jotka auttavat heitä kehittämään prosesseja ja tekemään strategisia päätöksiä liiketoiminnan kehittämiseksi.

2.2 Liiketoimintatiedon hallinnan historia ja kehitys

Liiketoimintatiedon hallinnan juuret ovat 1970-luvun lopulla kehitetyissä päätöksenteon tukiteknologioissa (Decision Support Technologies). Termin ”Business Intelligence” yleisti markkinatutkimusyhtiö Gartner 90-luvun vaihteessa. Tarkka vuosi on lähteistä riippuen joko 1989 tai 1993, joten ilmeisesti Gartner on tarkentanut määritelmäänsä. (Lawton, 2006, 14; Jagielska et al., 2006, 228)

Termiä ”Business Intelligence” on kuitenkin käytetty ensimmäisen kerran jo vuonna 1958, jolloin IBM julkaisi H. P. Luhnin artikkelin nimeltä ”A Business Intelligence System”. Luhnin paperissa esitetyt konseptit ovat hämmästyttävän samanlaisia kuin nykypäivän tutkijoiden esittämät. Luhn kirjoittaa että tietoa syntyy ja sitä käytetään yhä nopeammalla tahdilla kun ihmisten toiminta laajenee. Samaan aikaan yritysten kasvu ja osastoituminen luo esteitä tiedonkululle. Lehn painotti että on tarve laajalle järjestelmälle, joka sovittaa yhteen kaikki organisaation tietoon liittyvät ongelmat. Hän kutsui järjestelmää liiketoimintatiedon hallintajärjestelmäksi (Business Intelligence System). Järjestelmän päämäärä on tarjota yksilön, ryhmän, osaston, divisioonan tai jopa suuremman yksikön toimintaa tukemaan soveltuvaa tietoa. (Jagielska et al., 2006, 228)

Nykymuotoisen liiketoimintatiedon hallinnan voidaan katsoa syntyneen teollisuudessa 80- ja 90-lukujen vaihteessa. Pää syy sen syntyyn oli johtajien vaatimus taloudellisesta ja tehokkaasta yritystiedon hallinnasta ja analysoinnista, jotta he pystyisivät paremmin ymmärtämään liiketoiminnan tilannetta ja saamaan tukea päätöksenteolle. (Golfarelli et al., 2004, 1).

Voidaankin sanoa että liiketoimintatiedon hallinnan perimmäinen tarkoitus on säilynyt samana jo lähes kymmeniä vuosia. Kuitenkin vasta tietokoneiden ja tietojärjestelmien yleistymisen myötä on ollut mahdollista ryhtyä rakentamaan jo aiemmin kuvaillun kaltaisia järjestelmiä. Nykypäivän tietojärjestelmien mahdollisuudet tarjoavat liiketoimintatiedon hallinnalle valtavasti enemmän mahdollisuuksia kuin vaikkapa kymmenen vuotta sitten käytössä olleet järjestelmät. Liiketoimintatiedon hallinnan määritelmä ja konseptit ovat kuitenkin säilyneet lähes muuttumattomina koko

kehityskaarensa aikana, joten Jagielskan ja kumppaneiden (2006) peräämä keskustelu olisi tarpeen jotta nykyaikaisten IT-järjestelmien tarjoamat mahdollisuudet saadaan täysimittaisesti hyödynnettyä myös liiketoimintatiedon hallinnassa.

2.3 Liiketoimintatiedon lähteet ja muodot

Liiketoimintatiedon lähteet voidaan jakaa kolmeen kategoriaan: sisäiset tietolähteet, henkilökohtainen data ja ulkoiset tietolähteet. Sisäisistä tietolähteistä saadaan dataa yrityksen työntekijöistä, tuotteista, palveluista ja prosesseista. Henkilökohtainen data on yrityksen henkilöstön tuottamaa dataa, joka voi sisältää esimerkiksi konsepteja, ajatuksia ja mielipiteitä. Ulkoisista tietolähteistä saatava data on kaikkea yrityksen ulkopuolelta, kuten Internetistä saatavilla olevaa dataa (Turban et al., 2004, 495).

Liiketoimintatiedon hallinnassa käytettävä data voi olla joko rakenteellista (structured data) tai ei-rakenteellista (unstructured data). Ei-rakenteellista dataa voidaan kuvata myös termillä puolirakenteellinen (semi-structured), koska kaikella datalla on yleensä jonkinlainen rakenne. Rakenteelliseksi kutsutaan dataa, jota saadaan yrityksen erilaisista tietojärjestelmistä kuten ERP (Enterprise Resource Planning) ja CRM (Customer Relationship Management). Ei-rakenteellista tai puolirakenteellista dataa ovat esimerkiksi kuvatiedostot, tekstitiedostot, sähköpostit ja www-sivut. Taulukossa 1 on listattu muutamia esimerkkejä puolirakenteellisesta datasta. Markkinatutkimusyhtiö Gartner on lisäksi arvioinut että vuonna 2003 30-40% tietotyöläisten työajasta kuluu erilaisen ei-rakenteellisen tiedon hallinnoimiseen. Vuonna 1997 tehdyssä vastaavassa tutkimuksessa luku oli vasta 20%. (Negash, 2004, 180-181)

liiketoimintaprosessit	kirjeet	puhelinkeskustelut	uutisryhmät
pikaviestinkeskustelut	markkinointimateriaali	esitykset	videotiedostot
sähköpostit	muistiot	raportit	www-sivut
piirrookset	elokuvat	tutkimukset	white paperit
kuvatiedostot	uutiset	taulukot	tekstitiedostot

Taulukko 1. Esimerkkejä puolirakenteellisesta datasta (Negash, 2004, 180)

Yritysten liiketoimintatiedosta vain noin 20% on tallennettuna rakenteellisessa muodossa erilaisissa tietokannoissa. Jäljelle jäävä 80% on piilossa ei-rakenteellisissa tai puolirakenteellisissa dokumenteissa. Tämä johtuu siitä, että paras muoto tiedon ja tietämyksen ilmaisemiseen on teksti. Esimerkiksi markkinatutkimustulokset, tapaamispöytäkirjat, asiakasvalitukset tai kilpailijoiden mainokset ovat yleensä tallennettu tekstimuodossa. (Tseng & Chou, 2006, 728)

Verkossa julkaistavien dokumenttien jatkuva lisääntyminen ja julkisten dokumenttien hallintaan tarkoitettujen ohjelmistojen yleistyminen on johtanut tilanteeseen, jossa tietotyöläisten ja johtajien täytyy viettää suuri osa työajastaan lukien Internetissä julkaistuja erityyppisiä sähköisiä dokumentteja. Tämä määrä dataa on jo ylittänyt ihmisen kyvyn käsitellä ja ymmärtää sen sisältöä ilman tehokkaita apuvälineitä. Seurauksena on tärkeiden tietojen hylkääminen ja joskus myös hyödyttömän tiedon käyttäminen päätöksenteon tukena. (Tseng & Chou, 2006, 728)

Liiketoimintatiedon lähteiden jatkuvasti kasvava määrä ja uusien tietomuotojen tai tietotyyppeiden lisääntyminen aiheuttaa suuria haasteita liiketoimintatiedon hallinnalle. Käytettävien työkalujen täytyy pystyä käsittelemään valtavia määriä tietoa, joka on hyvin erilaisissa muodoissa. Rakenteettoman tiedon käsittely koneellisesti on hyvin haastavaa, mutta koska suurin osa yritysten tiedosta on puolirakenteellista ja yritysten työntekijöiden työajasta suuri osa kuluu tämän tiedon läpikäyntiin ja analysointiin, on liiketoimintatiedon hallinnalla saavutettavissa huomattavia etuja jo tavallisen valkokaulustyöntekijän työtehoa ajatellen.

3 LIKETOIMINTATIEDON HALLINNAN TEKNOLOGIAT

Tässä luvussa käsitellään liiketoimintatiedon hallintaan liittyviä teknologioita. Tärkeimmät aiheeseen liittyvät teknologiat ovat tietovarastointi, OLAP ja tiedonlouhinta. Luvun tavoitteena on antaa lukijalle perustiedot kustakin teknologiasta ja selvittää niiden merkitystä liiketoimintatiedon hallinnalle.

3.1 Liiketoimintatiedonhallinnan arkkitehtuuri

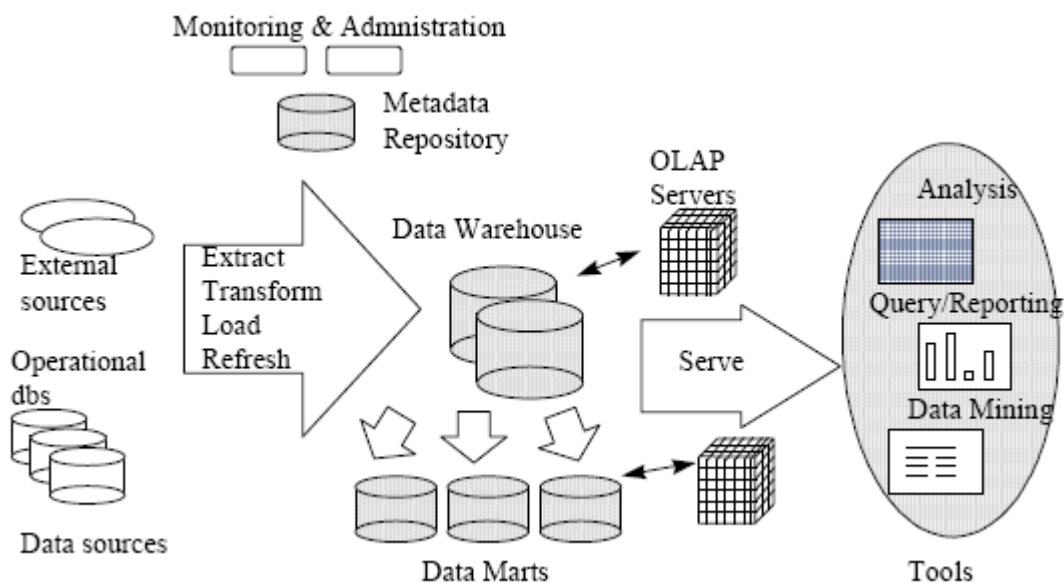
Liiketoimintatiedon hallinnan arkkitehtuuri rakentuu kolmesta teknologiakerroksesta. Azvine et al. (2005, 215) jaottelevat tarvittavat teknologiat lyhyesti seuraaviin kategorioihin:

- Tietovarastot (Data Warehouses) - keräävät ja tallentavat tietoa lukuisista lähteistä ja integroivat sen analyysia varten.
- Analyysityökalut – työkalut, kuten OLAP ja tiedonlouhinta, jotka ovat tarkoitettu tiedon analysointiin
- Raportointityökalut – muodostavat analysoituun tietoon perustuvia raportteja

Seuraavassa esittelen Kalakotan ja Robinsonin käyttämät määritelmät Liiketoimintatiedon hallinnan arkkitehtuurille. Myös he jakavat arkkitehtuurin kolmeen osaan, jotka vastaavat käytännössä Azvinen jakoa, mutta osat ovat hieman väljemmin määritellyt.

- Ydinteknologiat (Core Technologies) – tietovarasto (Data Warehouse)
- Mahdollistavat teknologiat (Enabling Technologies) – OLAP (Online Analytical Processing) sekä tiedonlouhinta (Data Mining)
- Liiketoimintatiedon hallinnan ratkaisut (BI Solutions) – ohjelmistoyhtiöiden ratkaisut liiketoimintatiedon hallintaan. Esimerkkeinä Microsoft Business Intelligence, Oracle Business Intelligence ja Cognos BI.

Kuvassa 1 on Chaudhurin ja Dayalin kaavio tietovarastoinnin arkkitehtuurista, mutta kuvassa on myös kaikki liiketoimintatiedon hallintaan liittyvät osa-alueet, joten sitä voi hyvin kutsua myös liiketoimintatiedon hallinnan arkkitehtuuriksi.



Kuva 1: Liiketoimintatiedon hallinnan arkkitehtuuri (Chaudhuri ja Dayal, 2006)

Arkkitehtuuri sisältää työkalut datan keräämiseen (Extract) useista operatiivisista tietokannoista (Operational dbs) sekä ulkoisista datalähteistä (External sources), työkalut kerätyn datan puhdistamiseen, muokkaamiseen (Transform) ja integroimiseen sekä käsitellyn datan tallentamiseen tietovarastoon (Load). Arkkitehtuuri sisältää myös työkalut tietovaraston ylläpitotoimenpiteisiin kuten ajastettuun tietojen päivittämiseen (Refresh) eri tietolähteistä sekä tietovaraston ”puhdistamiseen” siirtämällä harvemmin käytettyä tietoa esimerkiksi hitaammille arkistointipalvelimille. Päätietovaraston lisäksi järjestelmä voi sisältää useita yrityksen osastokohtaisia alatietovarastoja (Data Marts). Tietoa kaikissa tietovarastoissa hallinnoi yksi tai useampi tietovarastopalvelin, jotka muodostavat tiedosta moniulotteisen näkymän front-end-työkaluille kuten haku-, raportointi-, analyysi- ja tiedonlouhintatyökalut. Viimeiset arkkitehtuurin osat ovat tallennuspaikka metadatan tallentamiseen (Metadata repository) ja hallinointiin sekä työkalut tietovaraston valvontaan ja hallintaan (Monitoring & administration). (Chaudhuri ja Dayal, 2006)

3.2 Tietovarasto ja tietovarastointi

Liiketoimintatiedon hallinnan perusta on toimiva tietovarasto. Tietovarasto on järjestelmä, johon yrityksen eri tietokannoissa oleva data ja eri lähteistä saatava puolirakenteellinen tieto tallennetaan ja josta sitä tarjotaan erilaisten analysointityökalujen käyttöön.

Gardner (1998, 54) määrittelee tietovarastoinnin prosessiksi, jolla kootaan ja hallitaan useista lähteistä saatavaa dataa, ja jonka tarkoitus on tarjota yksityiskohtainen kuva koko liiketoiminnasta tai sen osasta. Chaudhuri ja Dayal (2006) kuvaavat tietovarastointia kokoelmana päätöksenteon tuen teknologioita, joiden tarkoitus on auttaa tietotyöläisiä (johtajat ja analyytikot) tekemään parempia ja nopeampia päätöksiä.

Voidaan siis sanoa että tietovarasto on järjestelmä, johon lukuisista lähteistä saatava tieto tallennetaan. Tietovarastointi puolestaan on prosessi, joka kattaa kaikki toimenpiteet, jotka liittyvät tietovarastoon. Näitä toimenpiteitä ovat esimerkiksi datan lataaminen yrityksen operatiivisista tietokannoista ja tämän datan puhdistaminen ja muokkaaminen tietovarastoon tallennettavaan muotoon.

Tietovaraston kehitys on seurausta W. Inmonin ja E. F. Coddin 90-luvun alussa tekemälle havainnolle, jonka mukaan operatiivisen tason tapahtumien käsittely (OLTP, On-line Transaction Processing) ja päätöksentekoa tukeva OLAP (On-line Analytical Processing) eivät voi tehokkaasti toimia samanaikaisesti samassa palvelinympäristössä kahdesta tiedon laatua heikentävästä syystä johtuen:

- Tiedon ominaispiirteet: OLTP tietokannat säilyttävät ajantasaista ja hyvin yksityiskohtaista dataa paikallisesti välitöntä operatiivista käyttöä varten. OLAP puolestaan käsittelee koostettua tietoa kattaen huomattavan suuren määrän eri tarkkuuksista ja erilaisilta ajanjaksoilta peräisin olevaa tietoa.
- Tapahtumien ominaispiirteet: OLTP:ssä korostuu hyvin pieneen tietokannan osaan kohdistuvien päivitystapahtumien tehokkuus kun taas OLAP vaatii pitkiä kyselyjä, jotka tutkivat kerralla hyvin suurta osaa tietokannasta. Näiden kahden samanaikainen yhdistäminen aiheuttaa hallintaongelmia. (Jarke & Vassiliou, 1997)

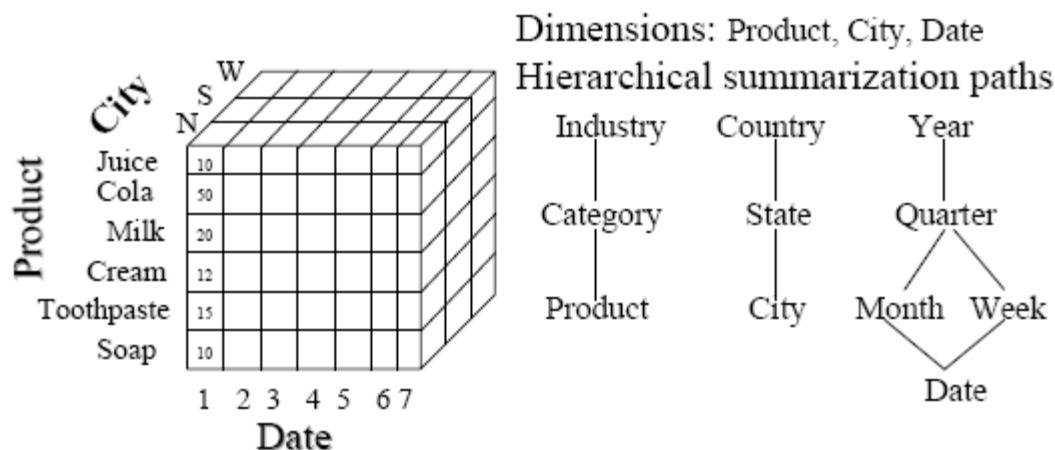
Tietovarasto on terminä hieman samanlainen kuin liiketoimintatiedon hallinta. Termille ei ole yhtä vallitsevaa määritelmää ja se voidaan käsittää hyvinkin eri tavoilla eri yhteyksissä. Yrityksissä tietovarastoja on rakennettu hyvinkin erilaisiin tarkoituksiin ja se osaltaan vaikeuttaa tietovaraston määrittelyä. Eräs viitatuimmista määritelmistä (Chaudhuri ja Dayal 2006; Akbay 2006; Lao ja Tang 2005) on kuitenkin William Inmonin ensimmäisen kerran vuonna 1992 (Lao ja Tang, 2005) esittämä määritelmä, jonka mukaan tietovarasto on ”a subject oriented, integrated, time-variant and non-volatile collection of data in support of management’s decision making process.” (Inmon, 2002, 31)

Lao & Tang (2005, 815) määrittelevät tietovarastoinnin datan aggregoinniksi eli tiivistämiseksi, jonka tarkoituksena on tukea johdon päätöksentekoprosesseja. Nykyään tietovarasto käsitetään yleisemmin ympäristönä kuin yksittäisenä tuotteena. Tietovarasto on sellaisten teknologioiden ja moduleiden yhdistelmä, mitkä integroivat operatiivista dataa päätöksenteolle tärkeään dataan pääsyn helpottamiseksi.

Tietovaraston neljä peruspiirrettä ovat (Lao & Tang, 2005, 815):

- aihekeskeisyys (subject oriented)
- datan integrointi (data integration)
- vakaus (non-volatile)
- aikasidonnainen (time-variant)

Tietovarasto, joka päivittyy operatiivisella tiedolla useista OLTP-tietokannoista, on myös OLAP-järjestelmän ydin. Tietovarasto järjestää tiedon käyttäen ulottuvuudellista mallintamista (dimensional modeling approach), joka luokittelee tiedon *mittoihin* ja *ulottuvuuksiin* (measures and dimensions). Mittoja ovat suureet kuten myynti tai kulut. Ulottuvuudet ovat mittojen eri perspektiivejä kuten tuote, sijainti ja aika. Ulottuvuudet järjestellään hierarkkisesti, joka mahdollistaa mittojen tarkastelemisen eri ulottuvuustasoilla (esimerkiksi liike -> kaupunki -> alue). Yhdistämällä mitat tiettyihin ulottuvuustasoihin luo tiedosta moniulotteisen näkymän, joka tunnetaan myös tietokuutiona (Data Cube). Kuvassa 2 on esimerkkikuva tietokuutiosta. (Thalhammer et al., 2001, 242)



Kuva 2: Tietokuutio (Chaudhuri ja Dayal, 2006)

Tyypillisesti tietovarastoja ylläpidetään erillään yrityksen operatiivisista tietokannoista. Tietovarasto tukee OLAP-tekniikkaa, jonka toiminnalliset ja suorituskyvylliset vaatimukset eroavat huomattavasti operatiivisten tietokantojen yhteydessä käytettävien OLTP vaatimuksista. OLTP sovellukset automatisoivat organisaatioiden päivittäistapahtumiin, kuten tilauksiin ja pankkitransaktioihin, liittyvää tietojen käsittelyä. Suoritettavat tehtävät ovat tarkasti määriteltyjä ja usein toistuvia. Tapahtumat vaativat tarkkaa ja ajantasaista dataa ja kohdistuvat yleensä muutamisiin osoitteisiin tietokannassa. Operatiiviset tietokannat ovat kooltaan sadoista megatavuista useisiin gigatavuihin. Tietokannan eheys ja palautettavuus ovat kriittisiä ja suoritettavien tapahtumien määrä on tärkein suorituskyvyn mittari. (Chaudhuri ja Dayal, 2006)

Tietovarastossa tiivistetty historiatieto on tärkeämpää kuin yksittäiset tarkat tietokentät. Koska tietovarastoissa säilytetään keskitettyä tietoa, monesti useista operatiivisista tietokannoista ja pitkältä ajanjaksolta kerättyä, ovat ne kooltaan huomattavasti suurempia kuin operatiiviset tietokannat. Suuryritysten tietovarastot ovat yleensä kooltaan sadoista gigatavuista useisiin teratavuihin. Tietovarastoihin kohdistuva haut ovat usein monimutkaisia ad-hoc -tyyppisiä hakuja, jotka voivat kohdistua miljooniin tietokenttiin kerralla ja niihin voi liittyä vaativia tietokantatoimenpiteitä kuten tietojen yhdistelyä ja koosteiden muodostamista. Tästä johtuen tietovarastoon kohdistuva kuormitus on hyvin suurta ja hakujen vasteaika on tärkeämpi kriteeri kuin suoritettujen tapahtumien maksimimäärä. (Chaudhuri ja Dayal, 2006)

Metadata tietovarastoinnissa

Metadataa on yleisesti kuvattu dataksi datasta. Relaatiotietokannassa metadata on tietokannassa määriteltyjen kohteiden kuvaus. Tietovarastoinnissa metadata viittaa kaikkeen, mikä kuvaa kohdetta tietovarastossa. Kohteita ovat esimerkiksi taulut, sarakkeet, kyselyt, raportit, liiketoimintasäännöt ja muunnosalgoritmit. Näiden määrittelyjen ymmärtäminen on kriittistä tietovaraston kehittämisprosessissa ja siksi metadatan hallinnan tulisi kytkeytä tiukasti kaikkiin tietovaraston osa-alueisiin. Tietovarasto on hyödyllinen vain tarjotessaan kilpailuetua, eli vain silloin kun siihen tallennettua dataa voidaan hyödyntää haettaessa vastauksia niihin liiketoimintatietoa koskeviin kysymyksiin joita varten tietovarasto rakennettiin. Metadata on kartta tähän dataan ja kertoo tietovarastossa olevan datan sijainnin ja tarkoituksen. Tietovaraston täytyy myös sisältää lista sisältämistään kohteista. Loppukäyttäjä tekee hakuja tietovarastossa olevaan dataan tätä listaa apunaan käyttäen ja tällöin hakua suorittavan prosessin täytyy tietää missä tarvittava data tietovarastossa sijaitsee. (Gartner, 1998, 60)

3.3 OLAP (On-Line Analytical Processing)

OLAP Council (1995) on määritellyt antanut OLAPille seuraavan määritelmän. OLAP on ohjelmistoteknologian luokka, jonka avulla analyytikot ja johtajat voivat parantaa näkemystään liiketoimintatiedosta saamalla nopean, johdonmukaisen ja vuorovaikutteisen pääsyn useisiin mahdollisiin näkymiin tiedosta, joka on muokattu raakadatasta vastaamaan yrityksen oikeita ulottuvuuksia käyttäjän ymmärtämässä muodossa. OLAPin toiminnallisuutta kuvaa dynaaminen moniulotteinen analyysi keskitetystä liiketoimintatiedosta, millä tuetaan loppukäyttäjän analyyttisiä ja navigoivia toimenpiteitä kuten

- laskutoimenpiteitä ja mallinnusta ulottuvuuksien halki, hierarkioiden läpi ja/tai jäsenien halki
- trendianalyysia peräkkäisten ajanjaksojen ajalta

- osajoukkojen viipalointia (slice)
- porautumista (drill-down) syvemmälle koostamisen tasolle
- kiertämistä (rotation) uusien ulottuvuuksien välisten vertailujen muodostamiseksi

OLAPin päämäärä on tarjota käyttäjille mahdollisuus suorittaa monimutkaisia analyysejä tiedosta helposti. OLAP tarjoaa graafisia työkaluja, joilla tiedosta muodostetaan moniulotteinen näkymä ja joilla käyttäjät voivat visualisoida, koota ja analysoida tietoa ja tutkia toistuvia kuvioita ja trendejä. OLAP-analyysin tulos voidaan esittää käyttämällä vuorovaikutteista raportointia. Raportointityökalut auttavat käyttäjää luomaan raportteja, joiden pitäisi tarjota päätöksentekijöille tietämystä (Intelligence) yrityksestään. (Jagielska et al., 2006, 232)

OLAP-järjestelmille on laadittu useita määritelmiä ja sääntöjä, jotka järjestelmän tulisi täyttää ollakseen käyttökelpoinen OLAPin päämäärän mukaisessa käytössä. OLAPin kehittäjäksi siteerattu E. F. Codd on luonut omat sääntönsä, joita on 18 kappaletta. OLAP Report -niminen tutkimusorganisaatio, joka tutkii ja vertailee OLAP-järjestelmiä, on luonut omat viisi sääntöään. Sääntöjen tarkoitus on auttaa yrityksiä vertailemaan ja valitsemaan OLAP-järjestelmiä. Seuraavassa käydään läpi sekä Coddin säännöt että OLAP Reportin FASMI-testi.

3.3.1 Coddin määritelmät ja säännöt

1993 E. F. Codd et al. kirjoitti paperin, jossa kuvataan OLAPin 12 sääntöä. Coddin työtä on kritisoitu, koska sen rahoittaja oli silloinen Arbor Software, nykyinen Hyperion Solutions. Tämän takia paperia on pidetty enemmänkin rahoittajan esitteenä kuin vakavasti otettavana tieteellisenä julkaisuna. Codd myös päivitti sääntöjään vuonna 1995 ja lisäsi kuusi uutta sääntöä ja jakoi lisäksi kaikki 18 sääntöä neljään ryhmään ja nimesi ne ominaisuuksiksi. Coddin 18 sääntöä ovat seuraavat: (OLAP Report; Codd et al. 1993)

PERUSOMINAISUUDET

- F1 Moniulotteinen käsitteellinen näkymä tietoon (Multidimensional Conceptual View, alkuperäinen sääntö 1). Loppukäyttäjä näkee yrityksen moniulotteisena ja OLAP-järjestelmässä olevan datan tulee vastata näitä luonnollisia ulottuvuuksia, jotta käyttäjän on helppo käsitellä dataa.
- F2 Helppo käyttöliittymä (Intuitive Data Manipulation, alkuperäinen sääntö 10). Käyttöliittymän tulee mahdollistaa monimutkaisetkin operaatiot ilman tarvetta valikkotoimintojen käyttöön.
- F3 Tietojen saatavuus (Accessibility, alkuperäinen sääntö 3). Käyttäjälle tulee tarjota kaikki tarvittava tieto, mutta ei ylimääräistä, yhdessä näkymässä riippumatta siitä missä lähteessä tieto sijaitsee.
- F4 Eristävä vs. tulkitseva (Batch Extraction vs Interpretive, uusi sääntö). OLAP-järjestelmän tulisi tarjota *sekä* oma ”välitietovarasto” käsiteltävälle datalle *että* tosiaikainen pääsy ulkoiseen dataan.
- F5 OLAP-analyysimalli (OLAP Analysis Model, uusi sääntö). Coddin mukaan OLAP-järjestelmän tulisi tukea neljää analyysimallia, jotka ovat: kategorinen (Categorical), tulkinnallinen (Exegetical), pohtiva (Contemplative) ja kaavallinen (Formulaic). OLAP Report muotoilee samat neljä mallia seuraavasti: staattinen raportointi (Parameterized Static Reporting), viipalointi ja kuutiointi sekä porautuminen (Slicing and Dicing with Drill Down), ”entä jos?”-analyysi (”What if?” Analysis) sekä maalinetsintämallit (Goal Seeking Models)
- F6 Palvelinarkkitehtuuri (Client-server Architecture, alkuperäinen sääntö 5). OLAP-järjestelmän palvelimen tulee olla joustava, jotta siihen voidaan helposti kytkeytyä erilaisilla päätelaitteilla.
- F7 Läpinäkyvyys (Transparency, alkuperäinen sääntö 2). OLAP-järjestelmä ei saa vaikuttaa loppukäyttäjän työskentelyyn riippumatta siitä kuinka OLAP loppukäyttäjälle tarjotaan ja missä käytettävä tieto sijaitsee.
- F8 Monen käyttäjän tuki (Multi-user Support, alkuperäinen sääntö 8). OLAP-järjestelmän tulee tukea useaa yhtäaikaista käyttäjää ja huolehtia samalla tiedon eheydestä ja tietoturvasta.

ERITYISOMINAISUUDET

- F9 Normalisoimattoman datan käsittely (Treatment of Non-Normalized Data, uusi sääntö). OLAP-järjestelmässä tehdyt muutokset dataan eivät saa vaikuttaa dataan sen alkuperäisessä lähteessä.
- F10 OLAP-tulosten tallentaminen (Storing OLAP Results: Keeping Them Separate from Source Data, uusi sääntö). Kaksisuuntaista OLAP-järjestelmää ei saa implementoida suoraan tapahtumatietoon (Transaction Data).
- F11 Puuttuvien arvojen poistaminen (Extraction of Missing Values, uusi sääntö). Puuttuvat arvot täytyy voida erotella nolla-arvoista.
- F12 Puuttuvien arvojen käsittely (Treatment of Missing Values, uusi sääntö). OLAP-järjestelmän tulee hylätä kaikki puuttuvat arvot riippumatta niiden lähteistä.

RAPORTOINTIOMINAISUUDET

- F13 Joustava raportointi (Flexible Reporting, alkuperäinen sääntö 11). Raportointia varten käyttäjän tulee pystyä asettelemaan ulottuvuudet vapaasti halaamallaan tavalla.
- F14 Tasainen raportoinnin suorituskyky (Consistent Reporting Performance, alkuperäinen sääntö 4). Järjestelmän on tarjottava loppukäyttäjälle tasainen suorituskyky vaikka ulottuvuuksien määrä ja tietokannan koko kasvaa.
- F15 Harvan matriisin dynaaminen käsittely (Dynamic Sparse Matrix Handling, alkuperäinen sääntö 7). OLAP-järjestelmän tulee mukautua matriisin koon ja tiheyden suhteen, jotta harva matriisi ei vie turhaa tallennustilaa eikä hidasta järjestelmän toimintaa.

ULOTTUVUUDENHALLINTA

- F16 Yleinen moniulotteisuus (Generic Multidimensionality, alkuperäinen sääntö 6). Datan ulottuvuuksien täytyy olla yhtäläisiä rakenteeltaan ja käsiteltävyydeltään.

- F17 Rajoittamaton määrä ulottuvuuksia ja tasoja (Unlimited Dimensions and Aggregation Levels, alkuperäinen sääntö 12). OLAP-järjestelmän tulee tarjota riittävästi tarkastelu-ulottuvuuksia (Coddin suositus 15) ja kunkin ulottuvuuden tulee tarjota käyttäjän tarvitsema määrä koostetasoja.
- F18 Rajoittamattomat ulottuvuuksien väliset operaatiot (Unrestricted Cross-dimensional Operations, alkuperäinen sääntö 9). Ulottuvuuksien välillä tulee sallia kaikenlaiset laskutoimitukset. Yleisimmät operaatiot tulee tarjota valmiina, mutta käyttäjällä tulee olla mahdollisuus ohjelmoida omia laskukaavoja.

3.3.2 FASMI-testi

OLAP Report (OLAP Report) kritisoi Coddin sääntöjä niiden suuren määrän takia. 12 tai 18 sääntöä on liikaa muistettavaksi joten OLAP Report on tiivistänyt OLAPin määritelmän viiteen avainsanaan: Fast Analysis of Shared Multidimensional Information (FASMI). FASMI-testin avulla on helppo vertailla eri OLAP-järjestelmiä, koska viisi avainmääritelmää ovat helposti läpikäytävissä ja ne ovat tuoteriippumattomia. Viisi avainmääritelmää ovat seuraavat:

- Fast – nopeus, vasteaika. Järjestelmän on annettava käyttäjälle vastaus keskimäärin enintään viidessä sekunnissa. Järjestelmän vasteaika on tärkeää, koska käyttäjän joutuessa odottelemaan hän hermostuu ja analyysityön taso laskee.
- Analysis – analysointikyky. Järjestelmän tulee pystyä käsittelemään kaikki loppukäyttäjälle tärkeät toimenpiteet ja toimenpiteiden suorittamisen täytyy olla yksinkertaista ja helppoa. Järjestelmän tulee myös mahdollistaa omien analyysien luonti helpolla tavalla.
- Shared – jaettu. Järjestelmän tulee tukea useaa samanaikaista käyttäjää ja varmistaa samalla luottamuksellisen tiedon tietoturva.
- Multidimensional – moniulotteinen. Moniulotteisuus on OLAP Reportin määritelmän avainominaisuus. Järjestelmän tulee tarjota käyttäjälle

moniulotteinen käsitteellinen näkymä dataan mukaanlukien useat hierarkkiatasot.

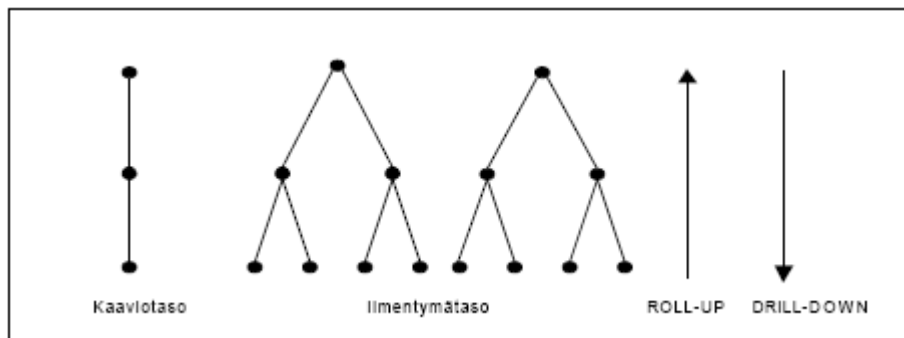
- Information – tieto. Järjestelmän tulee kyetä hallitsemaan kaikki tarvittava tieto riippumatta sen määrästä tai sijainnista

3.3.3 OLAP-operaatiot

OLAP mahdollistaa useita operaatioita tietokuutiossa olevalle tiedolle. Alla on esitelty operaatioista tärkeimmät, joilla näkymää tietokuution tietoon voidaan hyvin monipuolisesti muokata. Näiden operaatioiden lisäksi OLAP-sisältää useita operaatioita, joilla arvojen paikkoja voidaan vaihtaa

Roll-Up ja Drill-Down eli pyöristäminen ja porautuminen

Roll-Up –operaatio on tietoja tiivistävä operaatio. Operaatiolla siirrytään hierarkkiatasolla ylemmille ja samalla karkeammille tasoille. Drill-Down –operaatio on roll-up –operaation vastakohta ja sen avulla siirrytään yksityiskohtaisemmalle hierarkkiatasolle. Kuvassa 3 esitetään tietojen yhdistelyn suunnat molemmissa operaatioissa. (Kanerva, 2003, 22-23)



Kuva 3: Roll-Up ja Drill-Down –operaatiot (Kanerva, 2003, 23)

Slice eli viipalointi

Slice-operaatiolla tietokuutiosta voidaan valita tarkasteltava osa. Operaatio poistaa kuutiosta yhden tai useamman ulottuvuuden, jolloin käyttäjän näkökulmasta lopputuloksena on kaksiulotteinen sivu. (Kanerva, 2003, 23)

Dice eli kuutiointi

Dice-operaatiolla kuutiosta voidaan rajata pienempi osajoukko eli toisin kuin viipalointi kuutiointi säilyttää alkuperäisen kuution ulottuvuudet ja muodostaa rajauksen mukaan pienemmän tietokuution. (Kanerva, 2003, 23)

Pivot eli kääntö

Pivot-operaatiolla alkuperäisen kuution näkymää voidaan muokata rajaamatta pois tietoja. Pivot-operaatiolla ulottuvuuksien paikkoja tai suuntia voidaan vaihdella ja muodostaa näin uudenlaisia näkymiä tietoon. (Kanerva, 2003, 23)

OLAP-järjestelmä on tietojärjestelmä, jonka pääasiallinen tehtävä on tarjota käyttäjälle moniulotteinen ja käsitteellinen näkymä liiketoimintatietoon. Tiedon esittämiseen käytetään reaali maailman ulottuvuuksia, jolloin loppukäyttäjän on helppo hahmottaa esi ulottuvuuksien merkitykset ja niiden väliset suhteet. OLAP-järjestelmän tulee tarjota myös helppokäyttöiset analyysityökalut ja OLAP-operaatioiden avulla käyttäjä voi muokata näkymää haluamukseen. Näkymästä voi poistaa kokonaisia ulottuvuuksia tai tarkasteluun voi rajata vain pienen osan koko tietokuution sisällöstä. OLAP-järjestelmän raportointityökaluilla halutunlainen näkymä voidaan tulostaa raportiksi, josta analysoitua ja tiivistettyä tietoa tarvitsevat henkilöt voivat helpommin hahmottaa oleellisia asioita.

3.4 Tiedonlouhinta

Tiedonlouhinta on kuvattu prosessiksi, jolla etsitään käyttökelpoisia, uusia ja paikkansapitäviä malleja suurista datamääristä. Mallit voivat olla monessa eri muodossa: jos-niin-säännöt (If-then rules), matemaattiset yhtälöt, klusterit, päätöspuut (Decision Trees) tai näkyvä muoto. Tiedonlouhinta on kahdentyyppistä, ennustavaa ja kuvaavaa. Ennustava tiedonlouhinta on toistuvien mallien etsintää tulevaisuuden trendien ja käytöksen (kuten myyntien ja kohtennetun markkinoinnin) ennustamiseksi. Kuvaavalla tiedonlouhinnalla pyritään etsimään ymmärrettäviä malleja paremman liiketoiminnallisen näkemyksen saavuttamiseksi. Tiedonlouhinnalla voidaan saada aikaan hienostuneempia analyysejä kuin OLAP-työkaluja käyttämällä. Liiketoimintatiedon hallintaa ja tiedonlouhinta hyödynnetään laajasti esimerkiksi asiakkuuksienhallinnassa (Customer Relationship Management), jossa tiedonlouhinnalla voidaan ymmärtää asiakkaita ja heidän tulevaa käyttäytymistään paremmin. Tiedonlouhinnan avulla voidaan rakentaa historiadatasta malleja, joiden avulla voidaan esimerkiksi tarkastella eri asiakassegmenttien käyttäytymistä tietynlaisiin mainoskampanjoihin. (Jagielska et al., 2006, 233)

4 LIKETOIMINTATIEDON HALLINNAN SOVELLUKSET

Liiketoimintatiedon hallintaan tarjottujen sovellusten kirjo on hyvin laaja. Kaikilla suurilla ohjelmistotaloilla on nykyään markkinoilla oma ratkaisunsa ja jokainen näistä integroituu tiiviisti tai vähemmän tiiviisti yrityksissä yleisiin ERP eli toiminnanohjausjärjestelmiin (Enterprise Resource Planning), CRM eli asiakkuuksienhallintajärjestelmiin (Customer Relationship Management) ja muihin operatiivisiin tietojärjestelmiin.

4.1 Liiketoimintatiedon hallinnan sovellusten osa-alueet

Liiketoimintatiedon hallintaan tarkoitettut sovellukset sisältävät monia ominaisuuksia. Tietyt vaatimukset täytyy kuitenkin täyttää, ennen kuin valtava määrä raakadataa voidaan muuttaa hyödylliseksi tiedoksi. Kalakota ja Robinson (2001, 360) erittelevät viisi osa-aluetta:

- Datan kerääminen ja organisointi (Data/content organization and collection)
- Analysointi ja jaottelu (Analysis and segmentation)
- Reaaliaikainen personointi (Real-time personalization)
- Jakelu, palaute ja vuorovaikutus (Broadcast, retrieval and interaction)
- Suorituskyvyn tarkkailu ja mittaaminen (Performance monitoring and measurement)

4.2 Liiketoimintatiedon hallinta ja muut tietojärjestelmät

Päätöksenteon tukijärjestelmät

Liiketoimintatiedon hallintajärjestelmiä pidetään yleisesti päätöksenteon tukijärjestelmien (Decision Support Systems) seuraavana sukupolvena.

Liiketoimintatiedon hallinta on datan hyödyntämistä päätöksenteon tukena. Se tarjoaa näkemystä, joka auttaa johtajia tekemään taktisia päätöksiä sekä luomaan tai parantamaan liiketoimintastrategiaa ja prosesseja kilpailuedun ja muiden tavoitteiden saavuttamiseksi. Tietovarastointi, OLAP ja tiedonlouhinta, liiketoimintatiedon hallinnan avainteknologiat, ovat nykyisin osa strategiaa, jolla kehitetään päätöksentekoa. IBM pohtii että liiketoimintatiedon hallinta käsittää lähestymistavat, tekniikat ja työkalut, jotka käyttävät liiketoimintatietoa tarjotakseen vastaukset kysymyksiin, joita johtajat tyypillisesti kysyvät tehdäkseen parempia päätöksiä. (Jagielska et al., 2006, 234)

Operatiiviset tietojärjestelmät

Liiketoimintatiedon hallintajärjestelmät liittyvät oleellisesti yritysten operatiivisiin tietojärjestelmiin. Suuri osa tietovarastoon tallennettavasta ja liiketoimintatiedon hallinnassa käytettävästä datasta ja tiedosta on peräisin erilaisista operatiivisista tietokannoista. Lähes kaikilla merkittävillä toiminnanohjausjärjestelmien ja muiden operatiivisten tietojärjestelmien tarjoajilla on nykyään valikoimissaan myös liiketoimintatiedon hallintaan kehitetyt sovellukset, joten tulevaisuudessa liiketoimintatiedon hallinnan voidaan olettaa integroituvan yhä syvemmälle yritysten tietojärjestelmiin. Tämä on loppukäyttäjän kannalta hyvä asia, koska suuri määrä erilaisia työkaluja ja työhön tarvittavia tietojärjestelmiä hankaloittaa työskentelyä. Tiedonhallintatyökalujen tulo yhä useamman työntekijän ulottuville tarjoaa uudenlaisia tietämyksenluontimahdollisuuksia.

5 JOHTOPÄÄTÖKSET JA YHTEENVETO

Korkean teknologian ja tietointensiivisillä aloilla toimivien yritysten liiketoiminta perustuu yhä syvemmin liiketoiminnalle tärkeä datan ja tiedon hallintaan. Valtavat käsiteltävät tietomäärät vievät jo nyt suuren osan tietotyöläisen työajasta ja oleellisen tiedon puute heikentää päätöksentekijöiden päätösten laatua. Helpottamalla ja tehostamalla työkseen tietoa käsittelevien analyysityötä voidaan oleellisesti tehostaa työskentelyä ja parantaa analyysijä tarvitsevien päättäjien mahdollisuuksia parempiin päätöksiin.

Myös liiketoimintatiedon muodot ja lähteet lisääntyvät jatkuvasti. Internet mahdollistaa jatkuvasti uudenlaisen tiedon syntymistä ja esimerkiksi viime vuosina räjähdysmäisesti kasvanut sosiaalisten yhteisöjen suosio kiinnostaa varmasti monia yrityksiä. Sosiaaliset yhteisöt ovat vahvoja kanavia esimerkiksi hyväksi havaitun tuotteen vertaismainonnalle mutta vastaavasti negatiiviset kokemukset leviävät myös salamannopeasti ympäri maailman. Erilaisissa yhteisöissä ja blogeissa piilee valtava potentiaali esimerkiksi tuotekehitystä ajatellen kun jopa tuhannet käyttäjät kerääntyvät vertailemaan kokemuksiaan uusista tuotteista.

Liiketoimintatiedon hallinta tarjoaa työkalut, joilla valtavia tietomääriä voidaan käsitellä ja joilla raakadatan seasta voidaan etsiä liiketoiminnalle tärkeää tietoa ja jalostaa tämä tieto jatkokäyttöä varten. Liiketoimintatiedon hallinnan tavoite on tukea johtajien työtä tarjoamalla heille analysoitua tietoa päätöksenteon tueksi.

Liiketoimintatiedon hallinta koostuu kolmesta pääelementistä. Kaiken perustana on tietovarasto, johon analysoitava data ja tieto tallennetaan erilaisista lähteistä. Lähteitä ovat yrityksen omat sisäiset tietojärjestelmät sekä kaikki ulkoiset lähteet, joista liiketoiminnalle oleellista dataa ja tietoa voidaan kerätä. Tietovarastoinnille oleellista on myös valintojen tekeminen. Kaikkea saatavilla olevaa tietoa ei voida tallentaa tietovarastoon vaan on suoritettava karsintaa. Tärkeää onkin pyrkiä miettimään juuri omaa liiketoimintaa koskevien keskeisimpien tietojen lähteet, jotta tietovarastoa ei turhaan rasiteta turhalla tiedolla.

Tietovarastoon tallennettun datan käsittelyyn käytetään kahta muuta liiketoimintatiedon hallinnan pääelementtiä, OLAPia ja tiedonlouhintaa. OLAP-työkalut mahdollistavat käyttäjille helpon tavan erilaisten analyysien ja raporttien luomiseen. Käyttäjä voi tarkastella tietoa erilaisissa näkymissä ja luoda omia näkymiä tarpeen mukaan. OLAPin tärkein ominaisuus on juuri sen tarjoama moniulotteinen ja käsitteellinen näkymä tietoon. Tämä tarkoittaa sitä että käyttäjä voi tarkastella ja analysoida tietovarastossa olevaa liiketoimintatietoa käyttäen tiedon luonnollisia ulottuvuuksia ja mittoja. Työkalut tarjoavat myös helpot raportointitoiminnot, joilla käyttäjä voi tulostaa ulos haluamansa näkymän useissa eri muodoissa.

Tiedonlouhintatyökalujen avulla yritykset pystyvät etsimään ja havaitsemaan toistuvia malleja hyvin suurista tietomääristä. Konkreettinen esimerkki tiedonlouhinnasta on suomalaisten kauppaketjujen bonus-korttijärjestelmät. Nämä järjestelmät tallentavat tarkasti jokaisen ostotapahtuman ja kauppaketjuilla on näin jatkuvasti kasvava ostokäyttäytymisen tutkimisen mahdollistava tietokanta josta tiedonlouhinnan avulla voidaan pyrkiä esimerkiksi havaitsemaan tulevia trendejä.

Liiketoimintatiedon hallinta on kehittynyt päätöksenteon tukijärjestelmistä ja molempia käytetään myös yhä rinnakkain. Kirjallisuudessa on myös pohdittu onko termi ”Business Intelligence” vain yksi ohimenevä trendi tai vain uusi nimitys päätöksenteon tukijärjestelmälle. Oleellisin ero näiden kahden välillä on se, että liiketoimintatiedon hallinta sisältää oleelliset tiedon analysointiin tarvittavat työkalut. DSS-järjestelmät esittävät käyttäjille oleellista tietoa, mutta niissä harvoin on työkaluja analyysien tekoon ja tietojen muokkaamiseen.

Liiketoimintatiedon hallinnan seuraavaksi sukupolveksi nimitetään usein suorituskyvyn johtamisjärjestelmiä (BPM – Business Performance Management tai CPM – Corporate Performance Management). Suorituskyvyn johtamisjärjestelmät mahdollistavat yritysten liiketoimintastrategian systemaattisen suunnittelun ja seurannan sekä erilaisten syy-seuraus-suhteiden analysoinnin. Suorituskyvyn johtamisjärjestelmät ja niiden yleistyminen on myös merkki siitä että tiedolla ja sen tehokkaalla hallinnalla ja hyvin suunnatulla käytöllä on yhä suurempi merkitys yritysten liiketoiminnassa. Ennallaan säilyy siten vain jo 50-luvulla havaittu ongelma eli ihminen luo tietoa enemmän kuin pystyy sitä ikinä käyttämään.

LÄHTEET

- Abukari K. & Jog V. (2002). Business intelligence tools, CMA Management April, 76, 2, sivut 45-46.
- Akbay S. (2006). Data Warehousing in Real Time. Business Intelligence Journal. Vol. 11, No. 1, sivut 22-28.
- Azvine B., Cui Z. ja Nauck D. D. (2005). Towards Real Time Business Intelligence, BT Technology Journal Vol. 23 No. 3, July, sivut 214-225.
- Chaudhuri S. ja Dayal U. (2006). An Overview of Data Warehousing and OLAP Technology.
- Codd E. F. et al (1993). Providing OLAP to User-Analysts: An IT Mandate. [viitattu 18.7.2008]. Saatavilla osoitteessa <http://www.informatik.uni-jena.de/dbis/lehre/ss2005/sem_dwh/lit/Cod93.pdf>
- Gardner S. R. (1998). Building the Data Warehouse. Communications of the ACM. September 1998/Vol. 41, No. 9.
- Golfarelli M., Rizzi S., Cella I. (2004). Beyond Data Warehousing: What's Next in Business Intelligence?, ACM DOLAP '04.
- Hannula M. (2005). Liiketoimintatiedon hallinta. Tietojohtamisen peruskurssin luentomateriaali.
- Inmon W. (2002). Building the Data Warehouse (3. painos). Wiley & Sons, Inc. 412 sivua.
- Jagielska I. et al. (2006). Business Intelligence Systems for Decision Support: Concepts, Processes and Practice. Monash University, Melbourne Australia, sivut 227-240.
- Jarke M. & Vassiliou Y. (1997). Data Warehouse Quality: A Review of the DWQ Project. 2nd Conference on Information Quality. Massachusetts Institute of Technology, Cambridge 1997.

Kalakota R. ja Robinson M. (2001). e-Business 2.0: Roadmap for success 2nd edition, Addison-Wesley information technology series, 520 sivua.

Kanerva K. (2003). Lähestymistapoja OLAP-kieliin. Tampereen yliopiston tietojenkäsittelytieteen laitos, Pro Gradu –tutkielma. Saatavilla [www-osoitteessa <http://www.cs.uta.fi/research/theses/masters/Kanerva_Kaarlo.pdf>](http://www.cs.uta.fi/research/theses/masters/Kanerva_Kaarlo.pdf)

Lao G. & Tang Y. (2005). The Application of Data Warehousing in E-Business Environment and Case Study, ICEC '05, August 15-15, sivut 815-817.

Lawton G. (2006). Making Business Intelligence More Useful, Computer September, sivut 14-16.

OLAP Council (1995). OLAP and OLAP Server Definitions [online]. OLAP Council [viitattu 15.7.2008]. Saatavilla [www-osoitteessa <http://www.olapcouncil.org/research/glossaryly.htm>](http://www.olapcouncil.org/research/glossaryly.htm)

OLAP Report. The OLAP Report [online]. [viitattu 18.7.2008]. Saatavilla [www-osoitteessa <http://www.olapreport.com>](http://www.olapreport.com)

Thalhammer T., Schrefl M. ja Mohania M. (2001). Active data warehouses: complementing OLAP with analysis rules. Data & Knowledge Engineering 39, sivut 241-269.

Tseng F. S.C. & Chou A. Y.H. (2006). The concept of document warehousing for multi-dimensional modeling of textual-based business intelligence. Decision Support Systems 42, sivut 727-744.

Turban E., McLean E. ja Wetherbe J. (2004). Information Technology for Management: Transforming Organizations in the Digital Economy 4th edition, John Wiley & Sons, Inc., 731 sivua.