

Tietovarastointi (Data Warehousing) Finnairissa

*Urpo Löfbacka,
Kehityspäällikkö
Finnair Oyj*

Itsenäiskäyttö

Finnairissa tietovarastointi aloitettiin vuonna 1990. Tuolloin toimintaa kutsuttiin itsenäiskäytöksi. Lähtökohta oli että atk-yksikkö ei tee raportteja, vaan liiketoiminnan tuntijat laativat helpokäyttöisillä välineillä tarvitsemansa raportit itse. Atk suunnittelee yhdessä liiketoiminnan tuntijoiden kanssa selkeät otoskannat ja automatisoi niiden muodostusproseduurit. Itsenäiskäyttö nähtiin ylivoimaiseksi verrattuna Cobol-eräohjelmointiin nopeutensa ja joustavuutensa vuoksi. Ad-hoc -raportin tai hieman edelliskerrasta poikkeavan tulosteen saaminen atk-yksikön kautta kesti yleensä muutaman päivän, tai viikkoja. Ensin täytyi löytää suunnittelija, joka tuntee järjestelmän, josta tiedot saadaan ja sen jälkeen hänelle oli kyettävä esittämään tarve riittävän tarkasti. Yleensä ensimmäisellä kerralla asiaa ei ymmärretty samalla tavalla ja aikaa kului taas.

Finnairissa itsenäiskäyttö aloitettiin taloushallintoon ja lentokonehuoltoon liittyvistä sovelluksista. Syy tähän oli se, että nuo tietojärjestelmät sijaitsivat samassa laiteympäristössä, IBM-keskuskone, ja kummallekin alueelle oli 80-luvulla rakennettu massiiviset, operatiiviset tietojärjestelmät. Tukeen organisoitiin n. 1,5 henkilöä. Välineeksi valittiin IBM-keskuskoneella toimiva Focus, jonka valmistaja on Information Builders Inc. Keskuskoneen grafiikkaa emme edes harkinneet käyttäjä, vaan keskuskoneella möyhitty tieto siirrettiin työasemaan ja grafiikka tulostettiin taulukkolaskentaohjelmistolla. Käyttäjillä tällä konseptilla oli enimmillään n. 200. Käyttäjistä 20 osasi välineen läpikotaisin, 30 osasi laatia raporttiohjelman

alusta alkaen, 50 osasi tehdä raporttiohjelman malliohjelman muunnellen ja sadan käyttäjän Focus-aidot rajoittuivat pieniin parametri- ja ohjelmamuutoksiin.

Data Warehousing (tietovarastointi)

Liiketoiminnan asiantuntijoiden omaehtoista, otoksiin perustuvaa tietojen käsittelyä, alettiin 90-luvun puolivälissä kutsua Data Warehousingiksi. Data Warehousing -boomi oli kovimmillaan vuonna 1997. Miksi silloin? Tähän oli useita syitä, joista esitän tässä vain kolme: levytilan laskenut hinta, laiteympäristöjen pirstoutuminen sekä intranet.

Operatiiviset tietokannat ovat hyvin kompleksisia. Niiden rakenteella pyritään monimutkainen reaali maailma kuvaamaan mahdollisimman tarkkaan. Näiden järjestelmien tietokannoissa on samanaikaisesti läsnä kaikki mahdolliset näkökulmat tietoon. Operatiivisen järjestelmän relaatiokannassa jokainen tieto pyritään esittämään vain kertaalleen; tiedon monistus (redundanssi) on minimoitu normalisoinnilla. Otostietokannat sitä vastoin ovat kuhunkin käyttötarkoitukseen mahdollisimman selkeitä ja ymmärrettäviä. Tämän seurauksena sama tieto esiintyy useissa otoksissa. Ylläpito-ongelmaa tästä ei synny, koska otoskantoja ei päivitetä. Levytilaa kuitenkin tarvitaan paljon. 1990-luvulla levytilan hinnan lasku on jatkunut nopeana ja ensimmäisenä suuriin tietovarastoihin tarvittavan levykapasiteetin hinta alitti 'kipukynnyksen' Unix-ympäristöissä.

Vielä 1980-luvulla suuri osa yrityksen sovelluksista sijaitsi yhdessä laiteympäristössä, keskuskoneessa. Vuosikymmenen aikana on yrityksissä hankittu paljon valmisohjelmistoja ja osa

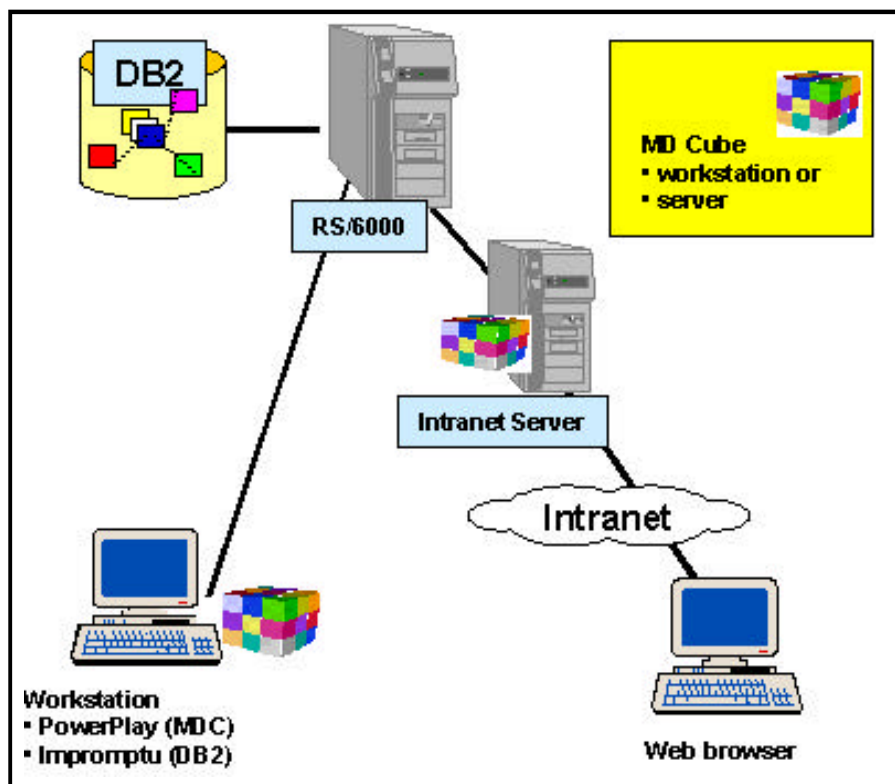
sovelluskehityksestäkin on siirtynyt Unix- ja NT-ympäristöihin. Keskuskoneiden aikaanakin tietojen yhdistäminen useasta sovelluksesta samaan raporttiin oli työlästä, mutta yrityksen laiteympäristöjen määrän lisääntyä räjähdysmäisesti, siitä on tullut käytännössä mahdotonta; ilman tietovarastointia.

Samanaikaisesti kun yritysten laiteympäristöt ovat 90-luvulla lisääntyneet, liiketoiminnan tarpeet ovat kehittyneet päinvastaiseen suuntaan. Yhä useamman tietojärjestelmän tietoja halutaan yhdistää samaan raporttiin tai analyysiin.

Suuri merkitys Data Warehousing -boomin voimakkuuteen, oli intranetin uusilla mahdollisuuksilla. Globaalisti toimivan yrityksen, kuten Finnair, yksi ongelma on tiedon tarjoaminen organisaatiolle siten, että kukin löytää oman työtehtävänsä kannalta merkitykselliset tiedot nopeasti. Ennen intranet-aikaa tiedon jakelussa käytettiin mm. paperiraportteja, levykkeitä tai faxia. Jos tiedoissa oli virhe, sama jakelu uudelleen selitysten kera. Intranetin vahvuuksia tiedon jakelussa ovat mm. tuttu käyttöliittymä, käytön helppous, alhaiset jakelukustannukset sekä mahdollisuus hyödyntää tietoja kaikkialta maailmasta ilman erillisiä ohjelmistoasennuksia. Tiedon jakelukanavana yrityksessä intranet on tällä hetkellä ylivoimainen.

Laiteympäristön ja DW-ohjelmistojen valinta

1997 asetimme tavoitteeksi asteittain luopua keskuskoneella tapahtuvasta tietovarastoinnista ja siirtä käyttäjän välineissä merkkipohjaisista graafisiin. Uusien välineiden valinnassa keskeisiä tavoitteitamme olivat seuraavat:



Kuva 1. Data Warehouse –arkkitehtuuri

- * perustietovaraston tulee olla relaatiotietokannassa
- * käyttäjävälineiden toimittajalla tulee olla selkeä visio ja pyrkimys kehittää tuotteitaan siihen suuntaan, että niitä voidaan käyttää myös selaimen avulla, ilman työasemakohtaisia ohjelma-asennuksia
- * välinetoimittajalla tulee olla riittävä markkinaosuus
- * tiedon siirrossa operatiivisista järjestelmistä tietovarastoon pyrimme yksinkertaisuuteen, jotta siirto eri ympäristöistä olisi mahdollisimman samankaltainen, peräkkäistiedosto on hyvä ratkaisu

Päädyimme kuvassa 1. esitettyyn Data Warehousing –ympäristöön.

Relaatiokannaksi valitsimme IBM DB2:n, joka Finnairissa oli laajassa käytössä jo entuudestaan. DB2:n UDB-versio oli joka suhteessa kilpailukykyinen muihin relaatiotietokannan hallintajärjestelmiin verrattuna. Käyttäjävälineiksi valitsimme Kanadalaisen Cognos:in tuotteet Impromptu ja PowerPlay. Impromptu on tarkoitettu relaatiokantojen käsittelyyn ja PowerPlay valmiiksi muodostettujen moniulotteisten kuutioiden käsittelyyn (MOLAP, Multidimensional OnLine Analytical

Processing). Raportointi on usein sitä, että lasketaan yhteen lukuja eri tekijöiden suhteen. Moniulotteinen kuutio on tiedosto, jossa summaukset on tehty valmiiksi useiden tekijöiden, dimensioiden, suhteen. Kuutio ei ole yleensä suurempi kuin tiedosto, josta se on muodostettu, mutta se on paljon nopeampi käytössä.

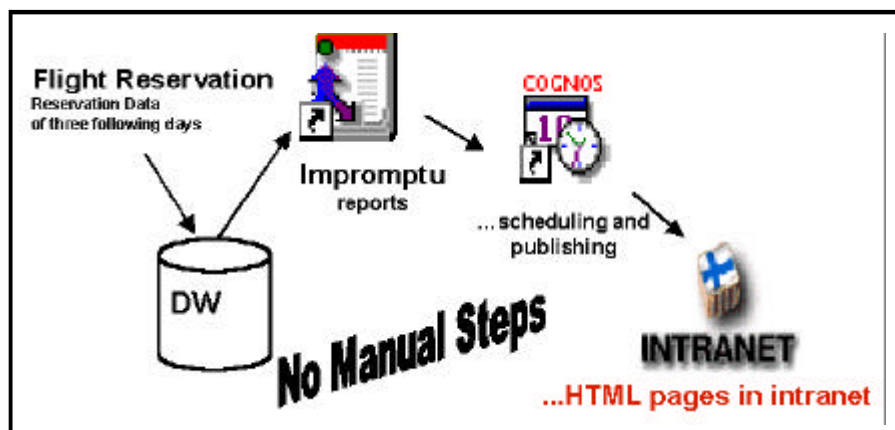
Sekä Impromptua että PowerPlay:tä tarvitaan ja nämä tuotteet kattavat hyvin perusraportointi- ja analyysitarpeet. Heti alkuvaiheessa päädyimme käyttämään myös PowerPlay –ohjelmiston Web-versiota, joka soveltuu erityisesti niille käyttäjille, jotka eivät tarvitse aivan kaikkia kuutionkäsittelyn ominaisuuksia tai joiden

koneisiin ei esim. fyysisten etäisyyksien vuoksi ole helppo tehdä työasemasennuksia. PowerPlay:n avulla on helppoa luoda intranetiin staattisia sivuja vain valitsemalla talletusmuodoksi html. Vaikka muodostuvat sivut ovat staattisia ne hyvin suunniteltuine valikkorakenteineen antavat 'dynaamisen' vaikutelman.

Nyt, 1.5 vuotta välinevalintojen jälkeen, näyttää edelleen siltä, että valitsimme meille parhaiten soveltuvat välineet. Välineiden merkitystä ei kuitenkaan tule liiaksi korostaa. Valittu välineistö ei todennäköisesti ole syy Data Warehousingin onnistumiseen tai epäonnistumiseen yrityksessä.

Vasta Data Warehousing –käyttöön-oton jälkeen tuli esiin konseptin suurin ja helpoimmin hyödynnettävä ominaisuus. Vanhoja keskuskonejärjestelmiä pidetään kankeina ja vaikeakäyttöisinä. Itse asiassa usein niiden operatiivinen osa on hyvinkin luotettavasti toimiva ja merkkipohjainen käyttöliittymä on tehokas erityisosaajille. Ongelma on usein se, että näihin järjestelmiin kerätty tieto ei ole tehokkaasti hyödynnettävissä muihin tarpeisiin. Yksi esimerkki tällaisesta järjestelmästä on lentopaikan varaus. Perusjärjestelmä toimii hyvin, mutta tieto ei ole helposti hyödynnettävissä esim. muissa järjestelmissä. Rakensimme tietovaraston yhteyteen mekanismin, jossa varaustieto siirretään joka yö ajastettuna tietovarastoon (AIX/DB2) jossa käynnistyy Impromptu-raporttiajo ja lopputulos siirtyy ajastettuna edelleen html-muodossa intranetiin (Kuva 2.).

Lentoasemalla lähtöselvityksestä vastaavat henkilöt näkevät varaustiedon ryhmiteltynä joko lennoittain tai esimer-



Kuva 2. Lennon varaustiedon(Unisys-keskuskone) jalostaminen ajastettuna intranet

LENTOPÄIVÄ	LÄHTÖAIKA	GOLD	SILVER	CLUB	VIP
11.11.1998	07:00-07:29	10	9	0	0
	07:30-07:44	18	9	0	0
	07:45-07:59	8	13	2	0
	08:00-08:14	20	31	1	1
	08:15-08:29	7	15	0	0
	08:30-08:39	10	8	0	0
	08:30-10:29	16	29	0	0
	11:30-12:29	3	9	0	0
	12:30-13:29	3	12	1	0
	13:30-14:29	7	8	0	0
	14:30-15:29	3	8	0	1
	15:30-15:59	4	2	0	0

Kuva 3. Asemapalvelun varaustietonäyttö (ennuste asemalle saapuvista matkustajista)

kiksi sen mukaan, milloin matkustajan arvioidaan saapuvan lentoasemalle (Kuva 3.).

Kuutioiden jakelu

Parhaat ratkaisut ovat usein yksinkertaisia. Tiesimme jo hyvin aikaisessa vaiheessa, että kaikille DW-asiakkaillemme ei riitä moniulotteisten kuutioiden käsittely selaimella tietoliikenneverkon yli. Osa haluaa käsitellä kuutiota työmatkoillaan tai kotonaan. Tällöin on järkevää siirtää kuutio käyttäjän työasemaan. Mutta kuinka järjestää uusien kuutioiden jakelu tehokkaasti. Ratkaisu oli yksinkertainen; kuutiot sijoitetaan intranettiin hakemistoksi, josta tietoa tarvitsevat lataavat ne omaan työasemaansa.

Tiedon puhdistaminen ja yhteismitallistaminen

Yksi tietovarastoinnin suurista eduista on, että se mahdollistaa useista eri lähteistä kerätyn tiedon yhdistelyn. Jotta tämä

olisi mahdollista, tiedon tulee olla virheetöntä, ristiriidatonta ja yhteismitallista. Yrityksen operatiivisissa järjestelmissä tieto ei juuri koskaan täytä näitä vaateita. Järjestelmät ovat syntyneet eri aikoina ja niiden lähtökohdat ovat kovin erilaiset. Esimerkiksi lennon numero esitetään eri järjestelmissä hieman toisistaan poikkeavalla tavalla; AY831, AY0831, 831, AY 831 jne. On valittava tapa, jolla tieto esitetään tietovarastossa.

Operatiivisiin järjestelmiin kätkeytyy aina virheitä, jotka eivät tule ilmi järjestelmää käytettäessä. Tällaisia ovat usein esim. 'blankko-avaimelliset tiedot' tai sama yritys esiintyy useaan kertaan kannassa hieman eri muodoissa. Näitä virheitä ei helposti huomaa näytöllä, mutta kun tieto siirretään tietovarastoon, mikään ei jää piiloon. Yleensä ei ole mahdollista korjata kaikkia operatiivisen järjestelmän virheitä, silloin kun tietovarastointi käynnistetään. Virheellistä tietoa ei kuitenkaan saa siirtää tietovarastoon. Hyvä periaate on se, että tiedot puhdistetaan ja

yhteismitallistetaan jo siinä vaiheessa kun ne siirretään operatiivisen järjestelmän tietokannasta peräkkäistiedostoon. Näin kaikki otoksen eri jalostusvaiheet ovat kaikki virheettömiä ja siten käyttökelpoisia.

Tietovarastoinnin tuki ja koulutus

Data Warehousing –tuki on Finnairissa järjestetty siten, että tehtävää hoitaa kaksi kaupallisen korkeakoulutuksen saanutta henkilöä. Hyvä kaupallinen koulutus yhdistettynä kykyyn ratkoa teknisiäkin ongelmia onkin ihanteellinen tähän työhön. Kaupallisen koulutuksen saaneelle tunnuslukujen käsittely on luontevaa ja tekniset ongelmat ratkeavat hyvin, mutta eivät muodostu pääasiaksi.

Tietovaraston sisältö ja välineet koulutetaan yhden päivän kurseilla. Koulutuksessa käymme koulutettavan tuntemaa aitoa tuotantoaineistoa. Tällä varmistetaan se, että koulutus saadaan välittömästi tukemaan liiketoimintaa. Ensimmäiset työssä tarvittavat raportit laaditaan jo koulutuksen aikana.

Hyvä johdanto tietovarastointiin on Ari Hovin kirja: Data Warehousing, Tietovarastointitekniikka, Suomen Atk-kustannus, 1997.

Urpo Löfbacka,
Kehityspäällikkö
Finnair Oyj
urpo.lofbacka@finnair.com