

SISU

PUBLIKATION 96:14

BEVAKNINGSRAPPORT – AUGUSTI 1996

Data Warehouse World 96

Stig Berild

SVENSKA INSTITUTET FÖR SYSTEMUTVECKLING

SISU

Innehållsförteckning

1. Inledning	1
2. Konferensdata	2
3. Intryck från konferensen	3
3.1 Trendbegrepp	3
3.2 Arkitekturer	4
3.3 Marts	5
3.4 OLAP	6
3.5 Kvalitet	8
3.6 Data Mining	8
3.7 Databas- och användarvolym	10
3.8 Detaljdata kontra summerade data	10
3.9 Erfarenheter	12
3.10 Införandestrategi	13
3.11 Metadata	13
3.12 Gränssnitt	15
3.13 Vad skiljer ett DW från en databas?	15
4. Intressant statistik från undersökning bland konferensdeltagarna	17
5. Uppsnappade synpunkter och tankvärdheter	18
5.1 Några synpunkter	18
5.2 Några tankvärdheter	19
6. Avrundning	20
Appendix	23
A1. Data Access inklusive replikering och kvalitetskontroll	23
A2. Repository, Meta Data	23
A3. Data Warehouse	24
A3.1 Vanliga Rdbms	24
A3.2 Specialiserade Data Warehouse-dbms	24
A4. Information Access	24
A4.1 Multidimensionella verktyg (OLAP) och EIS-verktyg	24
A4.2 Utsökningsverktyg	25
A5. Data Mining	25

1. Inledning

Många Data Warehouseföreträdare har i svepande termer talat om det begynnande informationssamhället med dess ökande globala konkurrens och allt nyckfullare kunder, om behovet av snabbanpassad produktionsapparat, om alla nya såväl vardagliga som strategiskt avgörande beslut som måste fattas. Mot hotbilden har ställts de unika möjligheter som kan yppa sig för de verksamheter som rätt förstår att ta till sig, anpassa och använda tillgängliga data för att generera kunskap och för att fatta välavvägda beslut i rätt tid.

Väl framme vid konkret realisering av dessa visioner har den raska vandringen övergått i ett desorienterat haltande, misslyckandena staplats på hög, besvikelsen brett ut sig.

Visserligen har de flesta verksamheter idag en uppsjö databaser i datahallar och på skrivbord. Ingen tvekan har rått om det allmänna behovet av att kunna nyttja dessa data för beslutsstöd på ett bättre sätt. Svårigheten har varit att veta hur. Här inkluderas såväl användningsaspekter som teknologi. Till yttermera visso har många under stor förundran snabbt kommit till klarhet om att denna verksamhetens gigantiska kunskapsreserv inte varit de härliga degklimpar som tillsammans med Data Warehouse-jästen skulle jäsa till oanad prakt och omfång – snarare ett antal torra, färdigstöpta och oförenliga limpor, bullar och kakor.

Historieperspektivet är alltså inte lysande. Samtidigt gör sig detta med informations-samhället alltmer påmint. Trycket på beslutsfattare ökar. Marginalerna mellan framgång och avgrund minskar. Riskbenägenheten för en Data Warehousesatsning ökar liksom möjligheterna att argumentera fram resurser för ändamålet. Störvulna satsningar på totalintegrerade Data Warehouse (DW) har ersatts med mer hanterbara, affärsområdesbaserade DW där chansen för "success stories" når en helt annan nivå. Efterfrågan på produkter ökar igen, vilket drar till sig mer investeringskapital som i sin tur genererar ett allt rikare produktutbud. Både positiva och negativa erfarenheter kan samlas in och ligga till grund för allt skickligare DW-hantering.

DW har alltså kommit ur sin svacka. Det är alldeles uppenbart att Data Warehouse nu är en företeelse i stark medvind. Inte minst visas det genom alla nya konferenser, kurser och seminarier som med förföriska lockrop erbjuder genvägar till informationslycka och framgång.

Föreliggande rapport redovisar intryck från den förmodligen i dagsläget mest framträdande och respekterade Data Warehousekonferensen "DCI's Data Warehouse World", Santa Clara, Kalifornien, 10-14 juni 1996 (DWW). Konferensen ges på flera platser per år men samlar enligt uppgift flest deltagare och utställare (kanske av naturliga skäl) i Silicon Valley.

2. Konferensdata

Konferensen samlade ca 850 deltagare. Många av dessa deltog dessutom i de tutorials (ett tiotal heldagsföreläsningar inom olika DW-relaterade teman) som gavs före och efter konferensen. Den parallellt pågående utställningen besöktes av ca 5000 personer. I båda fallen är det fråga om mer än en fördubbling mot föregående år. Intresset för DW växer med andra ord mycket snabbt. Över 200 utställare från ca 85 olika företag presenterade en uppsjö av olika produkter relaterade till DW. Produkter bl a för

- Databashantering med Data Warehouse-inriktning
- OLAP
- Data Mining
- Utsökning, rapportering av Data Warehouse-data (Användargränssnitt)
- Produkter för utsökning och transformation av operativa data
- Hårdvaror (parallelexekverare och minnen) för mycket stora databaser
- Metodstöd, övervakning, underhåll, ...

fanns representerade. Ett antal av dessa produkter finns uppräknade under några olika rubriker i ett appendix sist i rapporten. Syftet med appendixet är att dels ge en uppfattning om det rikhaltiga utbudet, dels ge uppslag till närmare efterforskning för dem som är i begrepp att investera.

Konferensen anordnades av en professionell arrangör vilket också märktes på genomförandet. Allt flöt mycket väloljat. Imponerande diger uppsättning presentationer – de flesta alldeles utmärkta – gjorde den vetgirige deltagaren snabbt villrådig. Mellan 8 och 14 parallella föredrag fanns hela tiden att välja mellan, alla med fängslande titlar. En frustrerande valsituation! Tre tunga volymer proceedings (en för varje dag) innehållande kopior av ljusbilder gav det mer permanenta minnesstödet.



Figur 1

Medaljens baksida är ett högt pris, ca 7000 kronor för konferens, därtill ca 2600 kronor för en heldagstutorial.

Nu över till intrycken.

3. Intryck från konferensen

3.1 Trendbegrepp

Det intressanta med de flesta IT-konferenser är att de ger en alldeles färsk spegling av vad som är "på gång", vad som ligger i den breda frontlinjen för stunden inom konferensens ämnesområde. Detta brukar manifesteras genom att alla plötsligt har samma trendbegrepp på läpparna. Alla utställare pekar följdriktigt på den egna produktens förträfflighet i dessa avseenden kompletterat med någon unik "knorr". DWW var i det avseendet inget undantag. Mycket kom att röra sig runt:

- Data Marts
- Data Mining
- Data Quality
- Terabytes of data (d v s våldsamt stora datamängder).

Bland mer udda men på sikt förmodligen intressanta och expansiva begrepp fanns:

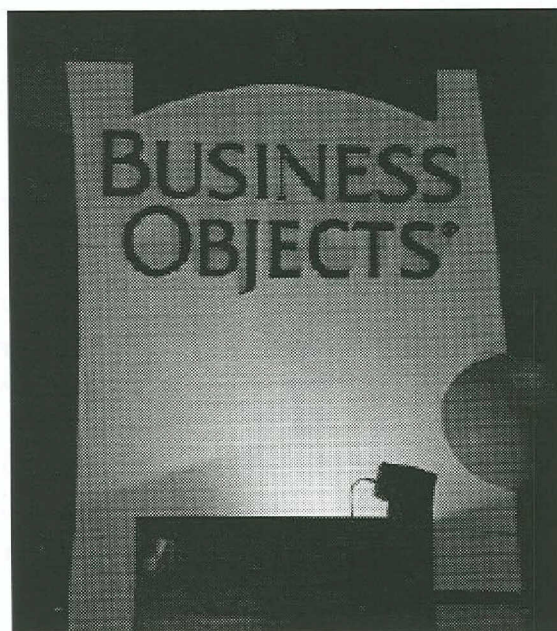
- Meta Data
- Meta Data Mining.

OLAP har varit i fokus under några år men nu blivit etablerad "vardagsmat" och föremål för en mer nyanserad debatt kring för- och nackdelar.

Självfallet hade konferensen även en hel del speglingar av de stora IT-strömningarna av typ:

- "Internet Access to the Data Warehouse"
- "Delivering the Data Warehouse with an Intranet"
- "The adoption of JAVA as a Mechanism for Data Access"
- "The World Wide Warehouse".

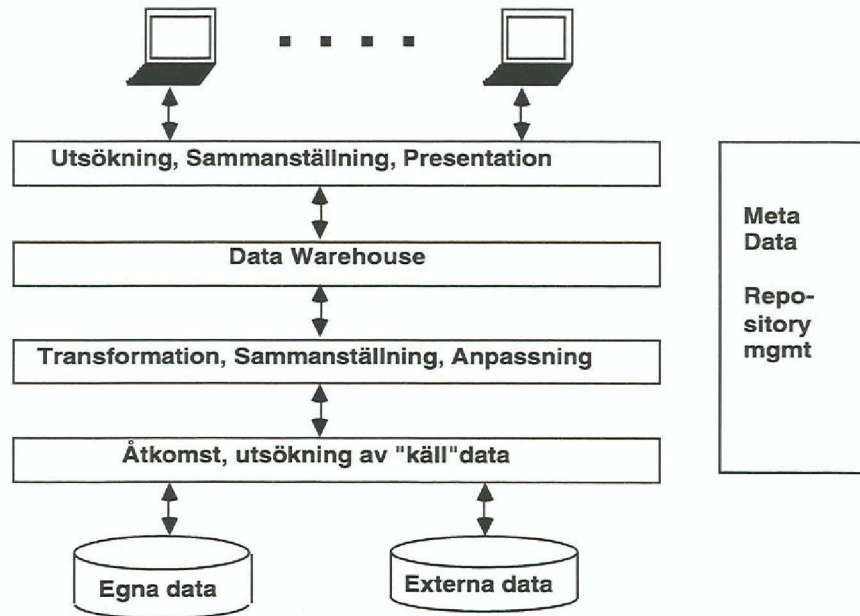
Givetvis kunde även "Business Objects" upptäckas i trängseln. Se figur 2.



Figur 2

3.2 Arkitekturer

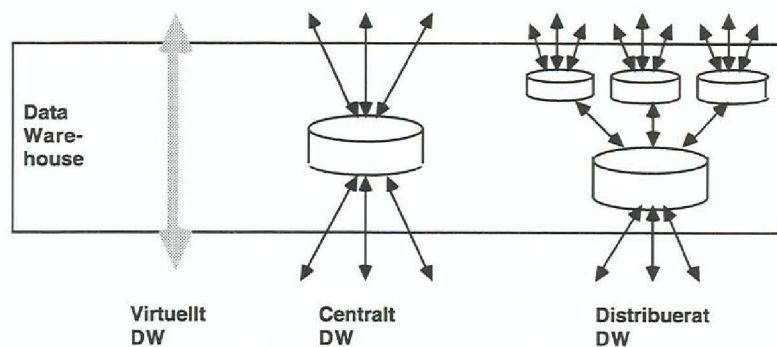
Stor samstämmighet tycks numer råda kring den typiska arkitekturen för en DW-miljö. Se figur 3. Samtidigt gör man en mycket klar distinktion mellan de "vanliga" databastillämpningarna i verksamheten (operational systems) och DW-tillämpningar (informational systems). I realiteten är det knappast möjligt att upprätta en kristallklar gränslinje mellan dem. Vad man vill poängtera är snarast att Informational systems i normalfallet jobbar med kopierade data och efter sina egna regler. Kontakten mellan systemtyperna inskränker sig, med vissa undantag, till leveranserna av källdata (grunddata, rådata).



Figur 3

Pilarna visar dubbelriktade snarare än enbart uppåtriktade för att indikera att DW-data kan komma att bli operativa data genom införlivande av felindikeringar, aggregeringar, m m av intresse för den operativa verksamheten.

Man brukar skilja på tre olika lösningar för Data Warehouse-rutan. Se figur 4.



Figur 4

Ett *Virtuellt DW* opererar direkt mot operativa data genom att utföra diverse utsökningar, sammanställningar, anpassningar i och för användarpresentation eller användarstyrd slutbearbetning inför presentation. Lösningen rekommenderades endast för verksamheter med infrekvent behov av DW-konsultation.

Centralt DW står för den konventionella tanken med en enda mer eller mindre integrerad datamassa (DW). Lösningen passar de verksamheter som antingen har en mycket välutvecklad DW-mognad och av den anledningen förmår operera med en integrerad lösning eller som opererar med en relativt begränsad datamodell, d v s i realiteten i form av ett avgränsat, affärsområdesspecifikt DW (Data Mart). *Distribuerat DW* innebär komplettering av ett centralt DW med Data Marts, antingen genom duplicering av samtliga data i det centrala DW till dessa Data Marts (teknikanpassning) eller genom duplicering av vissa uppgifter från centralt DW kompletterat med lokalt ämnesspecifika uppgifter i respektive Data Mart (affärsområdesanpassning).

Den distribuerade lösningen anses långsiktigt att föredra. Vad representerar då den nya trendbegreppet "Data Mart"?

3.3 Marts

Data Marts är ett relativt välbekant begrepp bland DW-folket. Det anses grovt stå för datamängder upp till 100 gigabytes (GB), d v s relativt små volymer i DW-perspektivet. Volymer mellan 100-500 GB kallas populärt för *Super Marts* och volymer däröver för *Hyper Marts*. För alla tre varianterna gäller samma syfte, endast volymen skiljer. Ett Mart syftar till att ge service till visst affärsområde, viss användarkategori, organisation, e dyl snarare än till verksamheten som helhet.

Tidigare ansågs skapandet av ett Data Mart vara ett finurligt första steg mot en DW-lösning. Man kunde koncentrera sig på behoven inom ett affärsområde. Dessutom var risken för konflikter kring semantik m m mindre. Inom ett affärsområde används ju förmodligen i betydande utsträckning vedertagna begrepp, välkända för dem som jobbar där. Framförallt större företag redovisar dock klart negativa erfarenheter. Visserligen kan Data Marts skapas med betydligt mindre möda än ett gemensamt DW. Men effekten blir ett antal fristående Data Marts med vattentäta skott emellan dem, trots mer eller mindre överlappande information. Olika semantiska nyanser, förmodligen uttryckta i olika kodformat, gör samverkan omöjlig. Resultatet blir samma problem som man förut haft med många okoordinerade databaser, nu endast upplyft till en högre nivå.

Följdaktligen påpekade ett antal föredragshållare att integrering av datamodeller och regler på en verksamhetsövergripande nivå ska vara grundbulten för all DW-verksamhet. Data Marts existerar lämpligen med en delmängd av centralt DWs datamodell och med data kopierade från detta. Syftet är främst att erbjuda anpassade användargränssnitt, lokala behörighetsmekanismer, aggregeringar, m m. Möjligtvis kan vissa kompletterande lokala data accepteras.

I teorin är detta förhållningssätt säkert korrekt. I verkligheten gäller en komplexitet och dynamik som tillsammans med skiftande krav inte tillåter en inledande utveckling av ett totalintegrerat DW ur vilket olika Data Marts senare genereras.

En mer nyanserad princip förs därför fram som alternativ lösning. DW innehåller här en gemensam kärna som respektive Data Mart sedan expanderar med att inkludera specialiseringar och områdesspecifika data. Problemet är bara att med säkerhet veta vad som är områdesspecifikt och vad som eventuellt kan visa sig vara överlappningar med omgivningen. En aktiv samarbetsorganisation med företrädare från alla Data Marts borde kunna erbjuda både kontroll och inspiration samt vara det reella beslutsorganet för lämpligt innehåll i det centrala DW.

Ytterligare andra vidhåller att det trots allt är vettigt att börja med Data Marts efter egna lokala förutsättningar och utan tyngande "byråkrati". Sannolikheten för ett lyckat resultat blir mycket större. Även en "success story" i det lilla är en inspiration till, och många gånger en förutsättning för fortsatt DW-användning och expansion.

Varje verksamhet måste givetvis utvärdera lämplig strategi baserad på egna förutsättningar.

Ett Data Mart är tills vidare ett enkel, tilltalande ansats på en rimligt hanterbar ambitionsnivå men med en ännu otydlig roll i förhållande till en övergripande DW-lösning.

3.4 OLAP

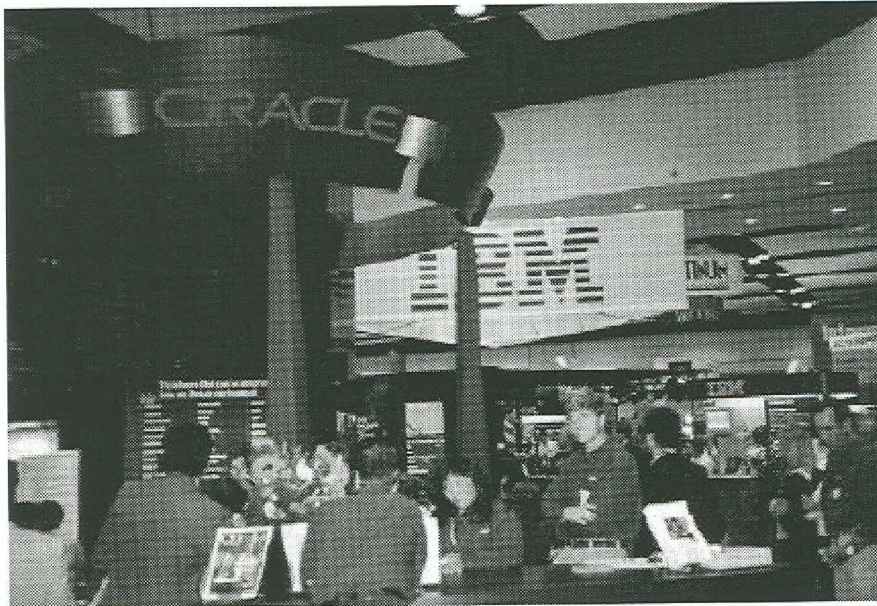
OLAP (On-Line Analytical Processing) har på senare tid enligt principen "kärt barn har många namn" expanderats till

- MOLAP (Multidimensional database based OLAP)
En integrerad OLAP-miljö inklusive erforderlig, ofta produktspecifik, databashantering.
- LOLAP (Local, desktop based OLAP)
OLAP-funktionalitet som kräver samverkan med separat dbms för nerladdning av de data man önskar operera på.
- ROLAP (Relational database based solution)
Komplettering av konventionella rdbms med OLAP-funktionalitet.

Produkter finns inom varje kategori. Den allmänna trenden tycks vara mot ett större utnyttjande av ROLAP. Tidigare har prestanda varit en hämsko. Nya optimeringsstrategier m m anses råda bot på de allvarligaste flaskhalsarna. Nya principer för att strukturera data likaså. Begrepp som "star schemas" och "snowflake schemas" representerar denna struktureringstrend. Å andra sidan riktas kritik mot dessa principer eftersom de tenderar optimera vissa kategorier frågor. Till förfång för – och kanske katastrof för – andra, senare tillkommande frågor.

Fördelen är ju att konventionella rdbms-databaser utnyttjas – inte slutna, produkt-specifika lösningar (som i MOLAP-fallet). Alla data ligger integrerade och tillgängliga genom SQL-gränssnittet, inte som en uppsättning separata "datakuber". Uppdateringar kan skötas på vanligt sätt och momentant återspeglas i olika aggregeringar. OLAP-databaser kan kräva en omfattande "refresh"-operation som en följd av en uppdatering. Nackdelen ligger i att man inte har samma möjlighet att optimera databasen för just OLAP-ändamål.

Knappast förvånande ser de stora rdbms-leverantörerna DW-marknaden som ett mycket angenämt och oproblematiskt område för expansion. De var väl företrädare på utställningsgolvet (figur 5). Kunderna investerar ogärna i nya eller produktspecifika teknologier om redan använd teknologi kan användas, inte minst om den är standardiserad och öppen.



Figur 5

En annan fördel påstås vara mindre minnesbehov jämfört med en multidimensionell lösning. En OLAP-databas sägs kunna bli flera gånger större än de grunddata den baserar sig på. Allt för att vara preparerad för snabba svar oavsett typ av fråga (attributvillkor, aggregeringsnivå). Å andra sidan kan säkerligen OLAP-företrädare ta fram exempel som visar på motsatsen.

I kompromissens tecken finns OLAP-verktyg som primärt arbetar med aggregerade data på olika nivåer och enligt önskat antal dimensioner, men som istället för att härbärgera grunddata, kan referera till dem i den DW eller operativa databas som primärt har att hantera dem. Lösningen erbjuder "drill-down" ända till grundnivån samtidigt som volymen på OLAP-databasen begränsas, ibland till en bråkdel av vad som annars skulle krävts. Observera dock att denna egenskap kan vara förenad med avsevärda restriktioner.

OLAP-området är "hett" varför intensiv fortsatt utveckling kan förväntas. Sannolikt kommer man att göra en klarare distinktion mellan multidimensionell upplevelse av data i användargränssnittet och multidimensionell databasteknologi. Att användare för vissa behov finner det både tilltalande och produktivt att kunna uppleva och operera på framförallt numeriska data, strukturerade enligt en multidimensionell modell, står alldeles klart.

Därmed inte sagt att den bakomliggande databasteknologin måste vara multidimensionellt uppbyggd. Användaren är knappast intresserad av teknologin annat än av prestandaskäl. Tillämpning av modern så kallad parallell databasteknologi förväntas ge avsevärda prestandaförbättringar för konventionella rdbms. Under förutsättning att prestandaproblemen löses, kommer marknaden för specifika OLAP-databaser att snabbt krympa. Lokala klient-baserade OLAP-databaser kommer förmodligen framgent att fylla ett behov som stöd för multidimensionella gränssnitt för hantering av olika typer av aggregeringar.

Givet denna trend kommer allt starkare krav på ett enhetligt gränssnitt baserat på en multidimensionell modell att föras fram. Bl a har Pilot Software pekat på sitt MDSQL som en lämplig kandidat att utgå ifrån i ett standardiseringsarbete.

3.5 Kvalitet

Om användarna upplever otillräcklig kvalitet hos DW-data upphör snart användningen. Å andra sidan bör inte målsättningen vara kvalitet till varje pris. Att kvalitetshöja inkommande data till ett DW kan vara förenat med stora kostnader eftersom arbetet bara till liten del kan automatiseras. Att föra tillbaka bristerna till de ansvariga för källdatabaserna brukar snarare skapa irritation än rättarvilja. Den operativa miljön kan mycket väl ha funnit sätt att leva med bristerna, alternativt är de där inte lika kritiska som för DW-tillämpningen. Problemen kan dessutom vara en konsekvens av bristande överensstämmelse, missuppfattningar, osäkerheter, vid sammanförandet av data från olika källor – något som knappast källdatabaserna kan lastas för.

För vissa analyser kan ofullständigheter och inkorrekta uppgifter mycket väl accepteras medan de i andra situationer kan leda till katastrofala beslut. Med andra ord bör alltid strävan vara att åstadkomma en kvalitet som svarar mot behoven, varken mer (kostnads-skäl) eller mindre (trovärdighet). Vilket låter sig sägas men i praktiken är en mycket svår balansakt.

Redan enkla data som namnfält i exv kunddatabaser visar sig ofta vara en stor källa till problem. En och samma kund återfinns normalt som ett flertal poster på grund av att namnet används som nyckel och stavats olika vid olika inmatningstillfällen. Dels får man ingen samlad bild av kunden, dels fördyras exv kontaktskapande aktiviteter. Ta till exempel ett postorderföretag som skickar ut flera miljoner kataloger. Det kan tjäna stora pengar enbart i tryck- och utskickskostnad genom att kunna rensa bort dubletter. Redan 20% reduktion av 5 miljoner registrerade kunder innebär 1 miljon färre utskick utan några som helst negativa konsekvenser. En föredragshållare redovisade en minskning av en fil över försäkringssäljare från 300000 poster till 27000 poster efter rensning med hjälp av ett avancerat datorstöd! En annan hade i en databas funnit 127 olika sätt att namnge företaget AT&T.

Kvalitetshandlingen kring ett DW bör angripas inom en överordnad IT-strategi, inte i form av spontana ”nödutryckningar”.

3.6 Data Mining

Data Mining (DM) förs allt starkare fram som en central DW-funktionalitet. Det är just DM som ska ge de stora lyften, de stora intäkterna, de starka argumenten för DW överhuvudtaget. I allmän överoptimism framställs det som något som ”get the answers without asking the questions”. Mer moderata företrädare drar snarare paralleller med gulddetning, något som givetvis är synnerligen gångbart vid en konferens i Kalifornien. Gulddetaren har vissa verktyg (Data Mining verktyget) och sin klokskap (strategier, hypoteser) till sitt förfogande samt en inmutning (en uppsättning data i lämpliga populationer) att leta inom. Ett översiktligt letande ger upphov till viss kunskap (mönster, regler, statistik) med vilken sannolikhetshypoteser om var guld finns och hur det kan tas fram formuleras. Om någon hypotes verkar tillräckligt stark påbörjas letningen.

Data Mining står för olika strategier för kunskapsgenerering och för prognostisering. Bland ansatser brukar i sammanhanget nämnas

- Neurala nät – nätverk av sammankopplade noder kapabla att upptäcka mönster i datamassan.
- Beslutsträd – data grupperade med hjälp av If – Then-satser.
- Regelinduktion – skapande av ickehierarkiska villkorsmängder.

- Fuzzy Logic – simulerar mänskligt resonemang genom att se ”gråskalor”.
- Datavisualisering – presentation över flera variabler (dimensioner) i en och samma grafiska figur.

Ansatzerna används för att generera ny kunskap i form av nya tidigare okända fakta, mönster, regler och hypoteser. Dessa upptäckter kan sedan utnyttjas för att generera prognoser, förutsägelser, bedömningar utifrån nya, tillkommande uppgifter.

Upptäckter brukar redovisas i form av

- Associationer – kopplingar till samma händelse.

Ett vanligt exempel som nästan alla tycks referera till är ett inte självklart samband mellan köp av blöjor och öl på lördagar, som man fann vid Data Mining mot en konsumentdatabas. Vid närmare analys visade det sig att män i USA ofta har ansvar för småbarn under veckosluten (blöjor) samtidigt som TV visar de intressantaste idrottsprogrammen (öl). Att i konsekvensens namn smyga in några backar öl i närheten av blöjförpackningarna glädjer säkerligen både affärsinnehavare och manlig kund.

- Sekvenser – kopplingar över tidsdimension.

En inte alltför uppseendeväckande upptäckt kan vara att de som nyligen köpt hus snart därefter blir kunder i vitvaruaffärer.

En annan kan vara att X % av de höftopererade inom Y månader kommer att anmäla knäproblem i andra benet (på grund av sned belastning).

- Klustring – notering av mönster i data utan föregående sökkriterium.

Kanske går det att ta fram att X% av ungdomar i storstadsområden i snitt köper Y CD-skivor per köptillfälle.

Ur ett annat DW går det att påvisa att produkt av typen X, tillverkad vid fabrik Y har ovanligt många klagomål varje år i december under de senaste 10 åren.

- Udda uppgifter och samband.

I en redovisad tillämpning fann man att basebollstjärnan X under de senaste fem åren mycket sällan lyckats göra mål när motspelare Y var på banan – en kunskap som givetvis föranledde en taktisk omdisposition.

Klagomålen över försenad tidningsutdelning har de senaste två åren under första kvartalet visat en fördubbling i distrikt X. (Vid närmare undersökning visade det sig att snöröjningen senarelagts inom distriktet med givna konsekvenser för utdelningen snöiga dagar.)

Hypoteser genereras löpande under Data Mining-arbetet. De som visar sig tillräckligt hållbara presenteras för användaren. Dessa hypoteser bör kontrolleras, vilket knappast är meningsfullt mot den redan genomgångna datamassan. Därför brukar man dela in datamassan så att DM-verktyget sätts att arbeta mot ca 75% av den varefter de framgenererade hypotesernas relevans testas mot resterande 25%. Håller hypoteserna kan de fortsättningsvis användas på aktuella inkommande uppgifter.

I de flesta redovisade DM-exempel är man ute efter kunskap om kunder; vad, när, varför de (inte) köper, deras köpprofiler, förmodade reaktion på nya produkter, o s v. Säkerligen kommer många andra områden att penetreras framledes.

DM är inom DW-området fortfarande inne i sin första fas även om vissa DM-entusiaster gärna för fram att skillnaden mellan DW och DM kan jämföras med distinktionen mellan data och information. Osäkerhet är ännu stor kring vilken typ av DM-stöd en kunskapsarbetare vid ett DW anser sig behöva. En absolut förutsättning är ett DW med klarlagd, rimlig datakvalitet. Existerande DM-verktyg anses heller knappast användbara för mycket stora DW. Dessutom kräver de mycket kompetenta användare som i första hand förstår den studerade verksamheten men även verktygets funktionalitet. Det för-tjänar också nämnas att ett antal av de exempel och upptäckter som redovisades är av en dignitet som vem som helst med en gnutta sunt förnuft och en smula eftertanke kan fundera fram utan större möda (ta bl a husköp-vitvaruköpsexemplet ovan). Kanske behöver fler erfarenheter baserade på mer seriösa eller effektskapande **upptäckter** redovisas för att marknaden ska "tända" ordentligt?

Innan eget ställningstagande rekommenderas den presumtive kunden studera, bedömma andras användning och erfarenheter av DM, gärna inom samma tillämpningsområde, såvida inte behoven är mycket väldefinierade eller tidskritiska.

Antalet produkter kan förväntas växa kraftigt inom detta ännu obrutna fält. Marknaden bedöms till ca 200 miljoner dollar 1996.

3.7 Databas- och användarvolym

Mycket snabb ökning av DW-databasvolym kan påräknas. Att tala om gigabytes i databassammanhang är "ute". Terabytes är "inne". 1 terabyte = 1000 gigabytes = innehållet i ca 11000 meter normalstora böcker. Som exempel på teknologi visade CRAY sin CS6400 Enterprise Server med en lagringskapacitet om ca 10 terabytes, 64 processorer för parallell exekvering, bus-kapacitet om 2 gigabytes/sekund samt ett 16 gigabytes primärminne; allt packerat i två personhöga "lådor". Bl a har man testat den med en Oracle-databas bestående av 9.2 miljarder tuppler (poster) om 1.9 terabytes, förutom index, m m. Det nära förestående intåget av multimediebaserade data i DW-databaser kommer snart att förpassa dessa siffror till ett mycket beskedligt härad.

Mycket snabb ökning av antalet användare per DW kan också påräknas. Flera tillämpningar rapporterar över tusen användare. Internet/Intranet-baserad åtkomst samt nät-distribuerad programvara kan förväntas bli norm framöver. Antalet användare anses i allmänhet inte bjuda några praktiska problem. Inte heller användarmotivation, m m. Bekymret ligger just vid den praktiska hanteringen av programvaror, releaser, rättelser så att enhetlighet och konsekvens uppnås. Å andra sidan är detta ett generellt bekymmer i klient/servermiljö.

3.8 Detaljdata kontra summerade data

Många förknippar ett DW och beslutstödssystem med operationer på aggregerade data på olika nivåer. Beslutsfattare vill bara ha övergripande, enkelt tillrättalagd information. Visserligen vill man då och då kanske titta lite mer på en mer detaljerad aggregerings-nivå men att lagra grunddata i ett DW skulle, anses det, skapa orealistiskt stora och ohanterliga databaser.

Här tycks nu trenden ha vänt. Dels klarar dbms idag datavolymer som bara för ett par år sedan ansågs ointressant stora. På något sätt tycks för övrigt behoven konstant svara upp

mot de tak som dagsläget erbjuder. Mycket vill alltid ha mer. Även om vi begränsar oss till i huvudsak textbaserade, strukturerade data finns behov av att hantera terabyte av data. Redan stora grunddatabaser resulterar i mångfald större DW-volymer i och med att även historiska data ofta är relevant att spara.

Trots volymerna anses grunddata vara en vital informationskälla att spara. Aggregeringar suddar ut nyanser, dessutom svarar de bara upp mot fördefinierade behov. Starka belägg finns för att nya informationsbehov är en nästintill kontinuerlig företeelse – kanske speciellt inom DW-området. Verksamheter förändras intensivt. Kunskap är det viktigaste konkurrensmedlet. Denna kunskap kan inte baseras på endast aggregeringar, medelvärden enligt gamla mallar. Dessutom, Data Mining förutsätter i allmänhet tillgång till källdata att ”leta i”.

Rob Armstrong, en av personerna bakom den kända varuhuskedjan Wal-Mart's DW-satsning hade i sitt föredrag ett antal intressanta kommentarer i anslutning till detaljer kontra summeringar:

- Inte minst konsumtionsindustrin arbetar numer mycket medvetet med att kartlägga sina kunders behov och köpbeteenden, bl a genom att föra in olika typer av kundkort att användas i kontakten mellan säljare och kund. (Konsums MedMera-kort, IKEAs Family-kort, m fl är exempel på denna trend.) Genom att lära känna kunden – in i minsta detalj – kan säljaren dels erbjuda ett mer lockande sortiment, specialerbjudanden, m m dels med välgrundade hypoteser närma sig nya kunder eller kundkategorier. De företag som lyckas är de som förstår sina kunder.
- Enda anledningen till att lagra aggregeringar är prestandahänsyn. Enda anledningen till att inte lagra detaljerade uppgifter är begränsad lagringskapacitet.
- Att endast tillhandahålla översiktlig kundinformation ger begränsat värde. Det finns ingen ”average customer”. Alla är unika. Olika typer av gamla, nya, modifierade analyser måste kunna utgå från denna unika information. Aggregerade data ger bara indikationer, ofta med stor osäkerhet. Kan vara bra för att få uppslag men krav på kompletterande detaljer följer ”som ett brev på posten” för att ge belägg. Ett använt schablonuttryck som speglar synsättet är ”You can't get the egg out of the cake”, d v s om data misshandlas genom olika bearbetningssteg eller omförs under aggregeringar et c (cake) är det omöjligt att ta fram de ursprungliga faktiska uppgifterna (egg).
- DW ger i viss mån åter den kundkunskap i stordrift som den gamla lanthandlaren hade om sin lokala kundkrets.
- En beslutsfattare ställer normalt fem frågor kring ett visst problem; de tre första avser aggregerade uppgifter, de två sista detaljuppgifter.
- Grundprincipen bör alltid vara att börja med att ladda detaljdata. Generera och lagra aggregeringar bara om tillräckligt många efterfrågar samma aggregerade uppgifter, eftersom varje ny detaljuppgift kan generera behov av ett antal mer eller mindre tunga uppdateringar av aggregeringsdata. Beräkna annars varje gång behov uppstår. Ju mer sällan en användare behöver en uppgift desto mer villig är denne att acceptera långa svarstider.
- Hur lång tid tar det att bygga ett DW? Svar: ”forever”. Hur lång tid tar det att få ut något värde av ett DW? Svar: ca tre månader.

Eller som någon annan sade i försvar av grunddata: ”When to aggregate? Wait until it hurts!”

3.9 Erfarenheter

Erfarenheter av stora DW börjar redovisas. Bland lyckade realiseringar nämndes bl a

- *Union Pacific Railroad* (vinnare av 1996 års "Realware" award for best Data Warehouse in USA)
Har integrerat tusentals, tidigare separata, filer i ett DW som ger ca 900 användare direktåtkomst till kritisk verksamhetsinformation.
- *Bank of America*
Har sammanfört data om över 23 miljoner kunder från ett antal olika system. Det primära syftet är en mer målinriktad marknadsföring samt bättre kreditriskanalys, banktransaktionsbedömningar, värdepappersanalys.
- *Whirlpool* (vitvaror)
Har samlat information om över 15 milj kunder och ca 20 milj hushållsapparater under upp till 30 års tid. Primärt för riktad marknadsföring, försäljningsanalyser och produktdesign (baserad på bl a reparationshistoria och kundpreferenser).
- *Wal-Mart*
Är det kanske mest kända exemplet på lyckad DW-satsning. Wal-Mart säljer alla upptänkliga sorters konsumtionsprodukter i ett stort antal affärer och varuhus spridda över USA. I ett ca 5 terabytes DW samlas alla uppgifter kring kunder, produkter, försäljning, m m vid samtliga varuhus över en rullande 65-veckorsperiod. En ovärderlig kunskapskälla har byggts upp. Avgörande konkurrensfördelar har erhållits, något som också framgår av Wal-Marts lönsamhet och marknadsandel.

Vissa DW- och operativa uppgifter görs även tillgängliga för leverantörer och grossister så att ansvaret för påfyllnad av varor kan överföras till dem. Wal-Mart har på så vis kunnat delegera den våldsamma lager- och påfyllningshanteringen till ett tidigare led i affärskedjan. Därigenom har man mer eller mindre kunnat inskränka sig till att erbjuda utrymme för produktsortiment i sina varuhus och vara "gränssnittet" mot kunden. Kapitalbindning minimeras eftersom leverantörerna äger produkterna i varuhuset ända tills kunden köper dem. Helt avgörande för detta upplägg är ett synnerligen avancerat och välfungerande DW.
- *American Airlines*
Använder sig av DW bl a för lagra diverse uppgifter om sina ca 28 miljoner Frequent Flyer medlemmar. Syftet är bl a att effektivisera riktade marknadsföringsinsatser, analysera kundernas flygmönster, m m.
- *McKesson* (läkemedelsdistributör)
Har produkt- och kundinformation i ett ca 1 terabyte DW för att bedöma produktlönsamhet per marknadssegment, ge service till leverantörer, detaljister, sjukhus kring produkter och konsumenter, m m.
- *U.S. West* (telefonbolag)
Samlar information om sina 25 miljoner kunder, deras abonnemang, servicebehov, m m i ett ca 1 terabyte DW för att kunna identifiera nya marknader, tjänster och för riktad marknadsföring, för lönsamhetsanalyser per segment, för att ge onlinestöd till ca 500 reparatörer, m m.

3.10 Införandestrategi

Många DW-projekt misslyckas. Bland orsakerna återfinns:

- Okunskap kring syfte.
- Alltför hög ambitionsnivå.
- Brist på rimlig och rimligt bred baskompetens hos såväl realiserare som användare.
- Teknikstyrt istället för användarstyrt. "They need ...", "They don't need ...", "What they really mean is ...", "If we build a DW, the users will come", "It'd be a great warehouse if it weren't for the users", o s v.
- Avsaknad av en realistisk kostnads/intäktsanalys. Ska upprättas även om det kan vara svårt att sätta belopp på vissa intäktsposter som exv "bättre beslut".
- Bristande stöd, förståelse, uthållighet hos budgetbeviljande ledning och/eller tänkta användare.
- Ingen klar gräns mellan operativa och DW-data, inte heller klar ansvarsuppdelning.
- Inflexibel lösning. Användare vet först efter längre tids användning hur de vill kunna operera på DW-data. I realiteten omöjligt att analysera fram i förväg.

Alltför lite intresse ägnas åt problemanalys innan DW-produkter köps in och tas i bruk. Den totala investeringen i teknik, hantering, kunskapsuppbyggnad, m m är i allmänhet våldsamt. I det läget borde det vara naturligt att tänka igenom den egna verksamhetens behov och förutsättningar för DW innan lockropen från leverantörerna ringer alltför starkt i öronen. Bland frågor att besvara är:

- Vilka problem avses DW lösa?
- Var finns "förtjänsten" att hämta?
- Vilka är de tänkta användarna och varför?
- Finns erforderlig uppsättning grunddata?
- Är grunddata tekniskt tillgängligt med rimlig möda?
- Hur är kvaliteten hos grunddata?
- Finns tillräcklig slutanvändarkompetens, teknikkompetens?
- Kommer verksamhetens organisation, arbetsflöden, funktioner, m m att behöva anpassas? I så fall hur?

Råd som återkommer i de flesta föredrag:

- Ha en välgenomtänkt plan. Gör en realistisk bedömning om genomförbarhet.
- Starta med ett mindre, välavgränsat DW för ett välformulerat behov som av ansvariga anses angeläget. (Data Mart)
- Se till att få garanti för en testperiod om minst ett år – de positiva effekterna kommer successivt.
- Bevisa användbarhet, "lönsamhet". Lyckat resultat ger "ringar på vattnet".
- Gör realistiska kostnadsanalyser. Skapa inte heller orealistiska förväntningar.
- Se till att etablera ett positivt samarbete med och förståelse hos ansvariga för de operativa systemen.

3.11 Metadata

Metadata börjar föras fram som en central kunskapskälla och som en förutsättning för att ett DW överhuvudtaget ska kunna opereras på med rimlig grad av säkerhet och kvalitet. Dock omnämns det förvånansvärt sällan. Möjligen förekommer begreppet repository eller metadata i en del figurer men de flesta tycks inte ta någon större notis om problemområdets vidd och angelägenhet. Det är ju trots allt genom metadata som förståelse om bl a verksamheten, vilka data som finns, hur de genererats, vad de står för och varifrån de kommer, kan uppnås.

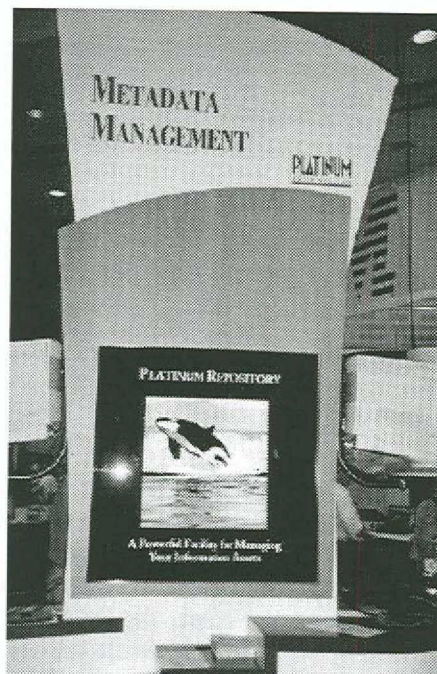
Källdatabasernas scheman är oftast den enda form av metadata tillgänglig för överföring till DW, i den mån den överhuvudtaget finns (alla källdatabaser är inte rdbms-baserade). Affärsrelaterad information såsom ändamål, regler, semantik, historik, behörighet, d v s den typ av uppgifter som en kunskapsarbetare mot ett DW önskar tillgång till, finns antingen inte alls eller ligger inbakad i databastillämpningarnas logik. Få diskuterar metadatastrukturens komplexitet i en realistisk miljö. Strukturen blir mycket omfattande om den dels ska representera summan av verksamhetens databasers scheman, dels DW-scheman, dels alla transformeringsregler dem emellan. För att inte tala om alla övriga affärsrelaterade uppgifter.

Metadata utgör även en viktig informationskälla vid administration av den totala DW-processen. Vem ser t ex till att ändringar i de operativa databastillämpningarna reflekteras in till DW-miljön? Erfarenhetsredovisningar pekar på avsevärda svårigheter att hålla DW korrelerat med förändringar i operativa affärsregler, struktur och innehåll. Inte minst hanteringen av historiska uppgifter kompliceras. Har data genererats under samma förutsättningar? Avser de exakt samma populationer? Gäller samma regler? Att jämföra momsbelopp mellan två år är skäligen ointressant om olika procentsatser gällde för respektive år och/eller om endast momsbelopp över x kronor registrerades under år x, alla belopp under år y, och beräknades för samtliga order from september för år z. Dessutom lagrades endast kreditorder i databasen under år y. Vem känner till de olika förutsättningarna och kan ta hänsyn till dem? Om inte, och det är det sannolika, uppstår koordinationsproblem. Vem löser dessa och hur?

Minst lika allvarligt är att även ett DW genomlöper revideringar, vidareutveckling m m som alla andra databaser. Detta får dock i allmänhet större effekter för ett DW eftersom historiska data, kanske etablerade enligt tidigare versioner av datamodellen, finns kvar under mycket lång tid. Även "metadata historia" bör med andra ord kunna hanteras och relateras till DW-innehållet.

Konstaterar vi därutöver att standard för strukturering av och gränssnitt till metadata saknas, blir det genast uppenbart vilka svårigheter den DW-ansvarige har att tackla vid realiseringen av ett metadatastöd.

Några repository-produkter ställdes ut, bl a den i figur 6.



Figur 6

3.12 Gränssnitt

Extrem användarvänlighet är ett måste för lyckosam exploatering av DW. Det betyder både effektiv och åskådlig presentation och intuitivt lättillgängliga funktioner. Vilket inte nödvändigtvis betyder bara några få enkla tangenter att trycka på. Beslutsfattare kan önska initiera avancerade beräkningar, kan vilja gräva vidare på detaljnivåer o s v. Det viktiga är att dessa operationer kan initieras i en enkel och rak dialog.

Dock; uppkomsten av en ny yrkeskategori, kunskapsarbetaren, anses oundviklig. Denne kommer att befinna sig mellan DW och beslutsfattaren med uppgift att ta fram önskade uppgifter, att analysera, bedriva Data Mining för att hitta ny information av vikt, m m. Kunskapsarbetaren måste vara flyhänt användare av tillgängliga programvaror, men i lika hög grad ha förståelse för verksamheten och beslutsfattarens behov av information. I arbetet ligger även att t ex bedöma osäkerheter, risker, variationer. Behörighet att arbeta över stora datamassor behövs och ställer kunskapsarbetaren i en särställning som kan kräva särskilda principer för övervakning och kontroll.

3.13 Vad skiljer ett DW från en databas?

Liksom andra modebegrepp inom IT har Data Warehouse många betydelser. Bl a

- | | |
|---------------------|--|
| Bill Inmon: | A subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management's decision-making process. |
| Herb Edelstein: | The consolidation of data from multiple sources into a query database. |
| Ken Orr: | A facility to provide easy access to quality, integrated enterprise data by both non-professional and professional end users. |
| VITAL Technologies: | An essential component of an information technology architecture dedicated to integration providing availability, sharability, and consistency within a flexible infrastructure. |

Med ett något annat fokus:

- | | |
|---------------|--|
| Ronald Swift: | Data Warehousing is a process , not a product or series of products. Data Warehousing is a technique for properly assembling and managing data from various sources for the simple purpose of answering questions and making decisions about your business that were not previously possible. |
|---------------|--|

För en ny betraktare av DW-området kommer snabbt en undran "Vad skiljer egentligen ett DW från ett vanligt dbms?". I båda fallen behöver man hantera stora datamängder i enlighet med något schema, måste en stor mängd samtidiga användare betjänas samt tycks SQL vara det tillämplade gränssnittsspråket. Några skiljelinjer finns dock:

- Framförallt olika tillämpningskrav och syften har resulterat i delvis olika funktionalitet. I en vanlig databastillämpning utförs normalt OLTP (On-Line Transaction Processing). En stor mängd uppdateringar utförs i allmänhet on-line. Korrekt hantering genom en avancerad transaktionsmekanism, behörighetskontroller, rollback-mekanismer, m m är oavvisliga krav.

- Ett DW uppdateras vanligtvis batchvis och med mycket stora volymer. Dessa data kommer från andra databaser, anpassade och strukturerade efter dessas förutsättningar. Därför måste normalt olika typer av formatkonverteringar, kontroller, aggregeringar, indexeringar, schemaanpassningar utföras "on the fly". Man talar också ofta om behovet av "to clean dirty source data" för att skapa en acceptabel kvalitet hos DW data.
- DW-användare utför primärt utsökningar. Utsökningarna tenderar bli betydligt mer komplexa och omfattande än inom OLTP. I vissa fall kan mycket stora datamängder behöva genomlöpas för att generera ett svar.
- Ofta är en DW databas betydligt större än en operativ databas, p g a att data sammanförts från många källor samt att historiska data lagras. Omfattande index för att snabba upp utsökningar tar också stort utrymme. Specialanpassad lagringsteknik kan behövas. Det är också ett känt fenomen att äldre DW data efterfrågas alltmer sällan, något som kan motivera uppdelning av datamassan i en del för snabb access och en annan billigare för långsammare access.
- En DW-användare arbetar ofta på varierande detaljnivåer. Ibland är höggradigt aggregerade data tillfyllest, ibland önskar man "drill down" för att se vilka underliggande data de aggregerade uppgifterna baserar sig på. Detta kan ske över ett antal aggregeringsnivåer ner till grunddata. Hantering av aggregerade data och deras dynamiska korrelation till underliggande nivåer är ett delikat problemområde i sig (prestandaavvägningar, konsistens, ...).
- Användare av konventionella databastillämpningar opererar normalt via formulär och/eller rapporter. DW-användare behöver lika normalt ytterligare ett antal användargränssnitt för ett effektivt analysarbete. Dit hör framförallt mångdimensionell funktionalitet typ OLAP. Oavsett om OLAP-verktyget arbetar med sin lokala datamängd (d v s gör en större utsökning och därefter arbetar lokalt med dessa data) eller är en integrerad funktion i ett DW (d v s varje bearbetning resulterar i en DW-operation) behövs olika typer av anpassningar.
- SQL saknar i dagsläget ett antal operationer som i DW-sammanhang är grundläggande och nödvändiga. Dit hör olika typer av statistiska operationer, tids-serieoperationer, operationer baserade på sorteringsordningar m m.

Även om listan ovan kan synas bestickande lång räknar de flesta med att den specifika DW-teknologin i ett längre perspektiv införlivas under den generella databashanterar-teknologin. Ur ett kundperspektiv finns heller inget egenvärde i att behöva arbeta med olika teknologier, standarder, användargränssnitt. Samtliga större dbms-leverantörer spelar också mycket aktiva roller på DW-spelplanen.

4. Intressant statistik från undersökning bland konferensdeltagarna

En högst ovetenskaplig, men dock, undersökning bland konferensdeltagarna gav bl a följande resultat:

- Data Marts är på stark frammarch. En fördubbling förväntas under det kommande året. Under 1997 beräknas ca hälften av alla DW-produkter vara avsedda för Data Marts.
- Antalet DW större än en terabyte kommer under det närmaste året att växa från 4% till 12%, d v s en tredubbling.
- Den starka frammarchen av DW kommer att gynna rdbms-leverantörerna eftersom de flesta DW kommer att vara rdbms-baserade.
- Antalet DW med fler än 500 användare kommer att öka från 4% till 12% av alla DW under nästa år – ytterligare en god inkomstkälla (licenser) för rdbms-leverantörer.
- Den totala DW-marknaden (mjukvara, hårdvara, service) beräknas öka från ca USD 2 miljarder 1995 till ca USD 8 miljarder 1998.
- Internet-anpassad DW-teknologi kommer att vara norm om ca 1.5 år.
- DW kretsar för närvarande nästan undantagslöst kring områdena
 - Finans
 - Marknadsföring
 - Försäljning
 - Kundkunskap
- På frågan

”What are the three (3) biggest challenges of implementing data warehousing projects within your organization?”

framkom följande svarsfrekvens för (endast den översta delen av listan visas):

35-40%	Data Quality
30-35%	Lagacy Data transformation/Scrubbing Managing User Expectations
20-25%	Managing Management Expectations Business Rule Analysis Managing Meta Data

5. Uppsnappade synpunkter och tankvärdheter

5.1 Några synpunkter

- Bekymret med DW är normalt att vissa parter måste ge mer än de får medan andra kan skörda frukterna. Samma sak gäller modellsamordning. Utifrån affärsområdesperspektivet är det sällan någon omedelbar nytta att samordna sina modeller med andra affärsområden för att uppnå övergripande enighet. Snarare kan det generera krav på oönskade anpassningar, m m. Alltså är det en mycket viktig uppgift för ledningen att formulera enkla och påtagliga argument för behovet av ett DW.
- Låt aldrig leverantörer styra inköp av produkter eller hur DW-verksamheten ska läggas upp. Endast kunden vet sina behov och prioriteringar. För att kunna agera i denna roll krävs dock kunskap. Vilket oavvisligen leder till att steg 1 alltid bör vara att inhämta gedigen kunskap och formulera klara behov.
- De allra flesta DW-exemplen kretsar i dagsläget kring konsumtionsvaruhandeln. Ur en amerikansk horisont med en enormt stor marknad spelar DW en om möjligt än mer central roll än för mer begränsade och/eller inte lika påtagligt konkurrensutsatta marknader.
- Erfarenheter visar att gränsen mellan operativa och DW-data är otydlig. I det dagliga operativa arbetet fattas stundligen beslut som i olika grad baseras på tillgänglig information. DW har av tradition i huvudsak syftat till att stödja strategiskt beslutsfattande. Gränserna kommer alltmer att suddas ut. Ta t ex en ordermottagare med ansvar att ge kundrabatter. Rabattsatsen grundar sig exv på kundens köphistoria, betalningsförmåga, tidigare klagomål och garantiarbeten, bedömd kundlojalitet hos kunder av samma kategori, beräknad efterfrågan under den närmaste tiden för de beställda varorna, o s v.
- Begreppet "Business Intelligence" är slagkraftigt även i DW-sammanhang. Visserligen skapar man sin "intelligens" primärt genom att bearbeta och dra slutsatser ur egna datamassor snarare än att gå ut och söka i externa källor men likafullt stärks verksamhetens beredskap, konkurrenskraft och överlevnadspotential. Mer inblandning av extern information i DW bedöms av många som naturlig och nödvändig i ett framtidsperspektiv.
- Värdet av DW blir synnerligen begränsat om det inte omgärdas av välgenomtänkt arkitektur och välfungerande administration, styrning, kontroll och ansvar.
- Tidigare har rent ekonomiska argument förts fram som syftet med DW, d v s att man förväntar sig tjäna pengar (Positive Return on investment) på DW-användning. En mycket klar omsvängning har på senare tid skett till att se DW som ett mycket verksamhetsredskap för "Strategic Business Initiatives", d v s med mer diffusa ekonomiska kvantifieringar och mer som en nödvändig ingrediens för avgörande beslutsfattande och för verksamhetens långsiktiga överlevnad. "Remember, Information is a competitive Weapon NOT a budget item". Syften bör ges i form av precisa, utvärderingsbara formuleringar som "öka omsättning i region X med minst 7% under det närmaste året" eller "öka återbesöksfrekvensen till 83% för registrerade kunder av kategori X", snarare än som "bättre beslut".

- Användning av DW kommer att generera gedigen övergripande kunskap om verksamhetens data, kvalitet och användning. Indirekt kommer denna kunskap att vara av stort värde vid omdesign, anpassning och integrering av framtida operativa system.
- Idag är i första hand kunskapsarbetare och beslutsfattare aktörer mot ett DW. På sikt kommer automatiserade Data Mining-operationer och olika typer av intelligenta agenter att i ökad utsträckning bli "DW-användare" i och för kunskapsinhämtning, övervakning, stöd, kontakter med operativa databaser, etc.
- Den stora kostnaden för DW ligger inte i den direkta DW-användningen utan i selektering, kontroll, transformation, rensning och överföring av operativa data till DW-databasen.
- DW är i grunden mycket enkelt, d v s är till för att stödja beslutsfattare med information hämtad från en redan befintlig datamassa. Många historiska, teknologiska bindningar, inflexibilitet och en teknikfokuserad oförståelse för information som kunskapskälla verkar hämmande. Osäker marknad och oklara behov har dessutom hämmat framväxten av bra verktyg. Detta håller dock på att vända i rask takt.

5.2 Några tänkvärdheter

- "Repository is the heart of DW. If you have multiple hearts you are in trouble. If you have no heart you are dead."
- På tal om diskussionen aggregerade data vs grunddata: "Summary Data -> Averaged Information -> Average decisions -> Average Performance".
- "Decisions are rarely either good or bad, they are based on good or bad information".
- "The right info in the right form at the right time" är en fjärran önskedröm för alla dem som idag arbetar mot alla massivt datautsprättande "legacy systems". Även nya system lämnar normalt onödigt mycket att önska i det avseendet. Användare – säg ifrån låt er inte teknikerstyras!
- "Metadata: what you wish you had in your data values".
- "Performance is like the weather – it changes every day – and someone's always complaining about it!".

6. Avrundning

DW har utvecklats från en otydlig idé, över en liten nischmarknad, vidare över en kostnadsbesparande IT-teknik för beslutsstöd, till en strategisk affärsprocessmekanism. Antalet DW expanderar snabbt både vad gäller antal tillämpningar, antal användare och databasstorlek.

Användning av DW kommer att gå över från att vara ett mer eller mindre magiskt, hemligt vapen för att vinna fördelar gentemot konkurrenter till ett alldeles nödvändigt redskap och stöd i många verksamhetsorienterade processer och för extern informationsservice.

Nya expertroller kommer att växa fram. Kunskapsarbetaren kommer att ges en klarare profil. En betydligt mer avancerad databasadministratörsroll kommer att krävas där teknikkompetens måste paras med en höggradig användarserviceprofil.

Fortfarande finns mycket att göra kring teknik för data mining, multimedidata och användargränssnitt.

Viktiga aspekter för en långsiktigt lyckosam DW-satsning är bl a

- datakvalitet
- tillgång på relevant metadata
- prestanda
- expanderbarhet dvs avseende datavolym, antalet användare, antalet källor, m m
- flexibilitet vad gäller datastruktur och möjlighet till spontana frågor
- användarvänlighet i presentationsgränssnitt och i hantering av mer komplex funktionalitet
- möjligheten till integration med eller åtminstone koppling till operativa data.

Internet/Intranet-baserad DW-åtkomst, browser-gränssnitt, applets för Data Mining ligger i närtid. Betydligt mer insamling av globalt tillgänglig verksamhetsextern information till det egna DW likaså. DW kommer successivt att integreras med övriga IT-miljöer i en verksamhet.

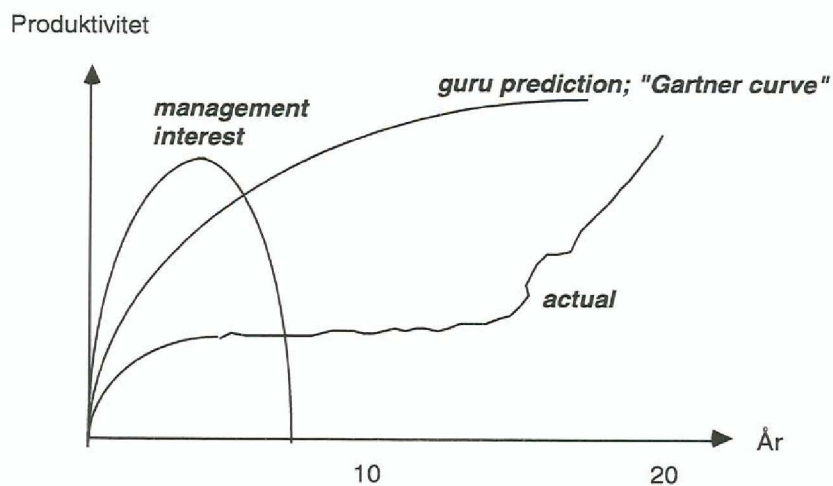
Både rdbms- och OLAP-produkter kommer att finna sina typiska tillämpningsområden. Inte minst kommer DW att vara en gyllene marknad för rdbms-leverantörer. Den snabba expansionen kommer att ställa höga krav på effektiv hantering av mycket stora databaser, inte minst i och med att DW kommer att popularas med multimediebaserade data. Dessutom kommer en rik flora av användargränssnitt och olika typer av avancerade analysstöd att växa fram. SQL kommer att i än högre grad inskränkas till ett internt databasgränssnitt.

Produkter kommer att bli mer heltäckande och integrerade. Mer avancerad funktionalitet kan förväntas.

En av de mest karismatiska DW-företrädarna, Ken Orr, har placerat in DW i ett teknikmognadsperspektiv där

- fas 1 innebär en uppbyggnad av problemförståelse
- fas 2 innebär en successiv förståelse för vilka lösningalternativ som är applicerbara
- fas 3 står för arbetet med att få det hela att fungera i verkligheten.

Han anser DW befinna sig någonstans i den senare delen av fas 2, något överlappande med de inledande stegen i fas 3. En mycket snabb fas 3-utveckling är att vänta, samtidigt som nya problem och lösningar dyker upp i de två tidigare faserna. Han konstaterar också de risker som alla nya teknologier är utsatta för, illustrerat i figur 7.



Figur 7

Prognosinstitut och "förståsigpåare" blåser upp förväntningar. De som ska fatta införandebeslut är på goda grunder skeptiska och tappar lätt intresset. I bästa fall överlever teknologin "riskzonen" och tar fart, dock efter betydligt längre tid än någon trott.

Den redovisade utvecklingstrenden har säkert hållit i historisk belysning. Dock bör varnas för att framförallt mer teknikbaserad utveckling numer tenderar att genomlöpa ett betydligt snabbare förlopp. DW-anpassad teknik kommer sannolikt att komma fram snabbt medan produktiv användning kommer att följa en betydligt svagare kurva.

Det övergripande intrycket från konferensen är en påtaglig entusiasm hos både användare och produktleverantörer, en stabil tilltro till DW-konceptet och en stark övertygelse om en snabbt expanderande DW-marknad. Potentialen anses enorm. Riskerna likaså, tyvärr.

Eller uttryckt med hjälp av ett stänk sommar:

Fjärden tycks ligga öppen och glittrande. Båtarna blir snabbt fler och allt större. Varning dock för förrådiska grund. Utprickning pågår men är ännu sporadisk. Endast den gode navigatören når målet.

Appendix; Några produkter

Produkter kommer och går i snabb takt. Leverantörer köps upp, ändrar inriktning, upphör. Nedan namngivna produkter visades på utställningen och/eller har omnämnts under konferensen. Listan är inte på något sätt fullständig, förmodligen inte heller fullständigt korrekt. Den ger förhoppningsvis ändå en uppfattning om DW-marknadens dagsläge.

För vidare information om produkterna hänvisas till leverantörernas web-sidor. De är i allmänhet föredömligt organiserade. Adressen är ofta www.<företaget>.com.

A1. Data Access inklusive replikering och kvalitetskontroll

Leverantör	Produkt
Apertus Technologies	Enterprise Integrator
Carleton Corporation	Passport
Computer Associates	LDM/XP
DBStar	DBStar Migration Architect
Evolutionary Technologies	Extract
Informatica	OpenBridge, PowerMart
Information Builders	Enterprise Copy Manager
Innovative Systems	Innovative Warehouse
Oracle	Open Gateway
Platinum Technology	InfoPump, InfoRefiner
Postalsoft	Address Correction and Encoding (ACE)
Praxis International	OmniReplicator
Prism Solutions	Warehouse Manager
QDB Solutions	QDB/Analyze, QDB/Connect
SAS Institute	SAS System
Smart Corporation	DB Workbench
Software AG	SourcePoint
Sybase	MDI
Vality Technology	Integrity Data Reengineering Tool

A2. Repository, Meta Data

Oracle CASE*Dictionary	Developer/2000
Brownstone DD/S	
CA/IDMS IDD	
RelTech DBXL	
Prism	Directory Manager
HP	IW Guide
IBM	Data Atlas, Data Guide/2
Rochade	
Platinum	Data Shopper

A3. Data Warehouse

A3.1 Vanliga Rdbms

Oracle
Sybase
DB2
Informix
CA-Ingres

A3.2 Specialiserade Data Warehouse-dbms

Teradata	
Tandem	
Red Brick Systems	Red Brick Warehouse
SAS Institute	SAS System
Praxis	
Fusion (IBI)	

A4. Information Access

A4.1 Multidimensionella verktyg (OLAP) och EIS-verktyg

Andyne Computing	Pablo
Arbor Software	Essbase
Business Objects	Mercury
Cognos	PowerPlay
Comshare	Commander OLAP
Holistic Systems	Holos
Hyperion	Enterprise
Information Builders	FOCUS/Fusion
Information Advantage	Decision Support Suite
Informix	MetaCube
IQ Software	
Kenan	Acumate ES!
MicroStrategy	DSS Agent
Oracle	Express
Pilot Software	LightShip Suite
Planning Sciences	Gentia
Platinum	Forest&Trees, InfoBeacon
SAS Institute	SAS System
Speedware	Media – EIS
Stanford Technology Group	MetaCube
Dimensional Insight	Cross Target

A4.2 Utsökningsverktyg

Andyne Computing
Borland
Brio Technology
Business Objects
Cognos
Crystal Computer Services
Dynamic Information Systems
IBM
Information Builders
Informix
INTERSOLV
IQ Software
MITI
Open Data Corporation
Oracle

Platinum
SAS Institute
Seagate

Sterling Software
Sybase

IBI WinViz
Computer Concepts dbExpress

GQL
ReportSmith
BrioQuery
Business Objects
Impromptu
Crystal Info and Reports
OMNIDEX
Intelligent Decision Server
FOCUS
NewEra ViewPoint
Q+E
IQ/Objects, IQ/SmartServer
SQR Workbench
FindOut!
Discoverer/2000, Data Query, Power
Objects
InfoQuery, InfoReports
SAS System
Crystal ReportsSoftware AG,
ESPERANT
Clear Access
Sybase IQ, OmniCONNECT,
InfoMaker

A5. Data Mining

Cross/Z International
DataMind
IBM
Information Harvesting
Information Discovery
NeuralWare
NeoVista
SAS Institute
SiliconGraphics
Software AG
Thinking Machines Corporation

Fractal Data Mining
DataMind
Intelligent Miner
Information Harvester
IDIS
NeuralWorks (IBMs?)
Decision Series
SAS System
MineSet
NETMAP
DARWIN