

INFORMATION WAREHOUSE

- VAD ÄR DET?

Stig Berild

SISU & SVERIGES TEKNISKA ATTACHÉER

©TRIAD Maj 1993

Innehåll

1. Problembakgrund 3

1.1 Problemsammanfattning 3

2. Behov 4

3. Lösningen: Information Warehouse 6

3.1 Principer 6

3.2 Sökningar över operativa data 7

3.3 Operationer på kopierade data 10

3.4 Synpunkter; operativa data 11

3.5 Synpunkter; kopierade data 11

3.6 Bedömning 12

4. Frågeställningar att beakta 12

4.1 Semantik 12

4.2 Ansvar 13

4.3 Funktion, strategi 13

4.4 Frågegränssnitt 15

5. Slutsatser 16

Litteraturförteckning 17

Bilaga: IBMs Information Warehouse Framework 19

Information Warehouse - vad är det?

Denna rapport består av två delar. Den första delen diskuterar företaget Information Warehouse generellt. Den andra delen är en bilaga som översiktligt beskriver IBMs nyligen lanserade Information Warehouse Framework. Information Warehouse är numer ett IBM-trademark. Som bakgrundsinformation har jag använt diverse nyligen publicerade artiklar från tidskrifter och nyhetsbrev, ett par konferensföredrag samt IBM-manualer.

Varmt tack går till Sten Andler vid IBM Santa Teresa Laboratoriet, San Jose som varit till stor hjälp med information, material och arrangemang av möten vid IBM.

Stig Berild

1. Problembakgrund

Datamassan växer snabbt i de flesta verksamheter bl a som en följd av att verksamheten blir mer komplex, att verksamheten och verksamhetens omgivning (kunder, myndigheter, leverantörer m m) ställer mer avancerade krav och att det blivit allt lättare att erhålla och hantera data (gränssnitt, dbms, datorprestanda, datorer-terminaler på varje arbetsställe m m).

Samtidigt ställs ökade krav på beslutsfattande och beslutsfattare. (Obs: Beslutsfattare är inte bara människor, det kan t ex lika gärna vara funktioner i en tillverkningsprocess.) Dessa ställer i sin tur ökade krav på mer heltäckande, väl utvalt, snabbt tillgängligt, överskådligt presenterat och högkvalitativt beslutsunderlag från de datasystem som hanterar datamassan.

Där börjar problemen:

- Data finns i allmänhet på många datortyper, under många olika operativsystem, under olika filsystem, databashanterare m m.
- Data på olika platser har olika grad av aktualitet, kvalitet, m m.
- Datas semantik är entydig inom viss tillämpning men normalt inte alls samordnad mellan tillämpningar, databaser etc.
- Vissa data finns inte tillgängliga som fristående element, de kan t ex vara inbakade i en textmassa, medan andra finns entydigt urskiljbara i en datastruktur.
- Data kan vara lagrade sekvensiellt i en äldre form av filer, vilket kan kräva en genomgång av en hel fil för att ta fram en uppgift. En sådan sekventiell genomgång är kostsam och stör dessutom produktionskörningarna.
- Han eller hon som behöver data måste begära dessa från olika ansvariga personer, som kanske måste programmera en utskriftsfunktion för att få fram det. Därefter måste erhållna data från olika platser integreras och koordineras innan informationen i realiteten kan börja användas.
- Omfattande data måste kunna väljas och sammanställas till en rimlig och överblickbar enhet för beslutsfattaren. Innehållet i denna sammanställning måste vara anpassat till i vilken slags situation det ska användas, typen av beslut, beslutsfattaren, typen av data o s v.
- I anslutning till detta har vi att ta hänsyn till hantering av härledda data och deras aktualitet gentemot dess grunddata. Även olika versioner, eller historik av data, bör kunna hållas samman. Hur vet man om refererade data är grunddata eller baserade på andra data av god eller dålig kvalitet? Känner vi till härledningsmekanismen för att kunna utröna kvaliteten?

1.1 Problemsammanfattning

Dessa förutsättningar kommer vi att få leva med, dessutom i allt större utsträckning enligt många uppfattning. Total samordning är i praktiken en omöjlighet i en alltmer föränderlig och komplex informationsvärld.

Man ställer ökande krav på att information ska vara tillgänglig samtidigt som denna information sprids alltmer. Detta motsatsförhållande har blivit ett problem, men det måste kunna överbryggas. I en manual från IBM (1) säger man:

"The modern enterprise, faced with competitive pressures, needs better quality information, which requires easier and faster access to increasing amounts of data; yet that data is often difficult to find, to access, and to understand."

2. Behov

Låt oss titta på behoven något mer konkret. I en verksamhet vill beslutsfattare och andra informationsarbetare kunna hantera data utan att behöva ta hänsyn till problem enligt ovan. De behöver kunna hantera data som en enhetlig massa och kunna referera till just de delar, som för tillfället bl a behövs för att:

- Göra summeringar och sammanställningar på verksamhetsnivå över vad som händer längre ner i organisationen.
- Göra trendanalyser.
- Få stöd i bedömningar av typen: Vad händer om...?
- Få svar på alla sorters spontana frågor över såväl sammanställda data som källdata.

Informationsarbetaren behöver också tillgång till data om data, ofta benämnt "meta-data". Detta för att rätt kunna ställa en fråga, analysera värde, kvalitet, fullständighet och tolkning av ett svar, för att kunna administrera och förädla data och databearbetning efter behov, för att kunna bedöma svarstider (data kan ju behöva hämtas från många källor) och kostnader m m.

Några frågor om metadata är:

- Vilka verksamhetsövergripande typer av data finns det och hur är de relaterade (med andra ord uppgifter som svarar mot datamassans schema)?
- Hur har olika typer av data skapats? En övergripande processmodell kan t ex ge indikation om beroendeförhållanden mellan datatyper, utsprung, duplikat, m m.
- Var finns data av olika typ? Hur är de semantiskt relaterade till den verksamhetsgemensamma uppfattningen?
- Hur kan man komma åt data och datatyper?
- Information om de härledningsregler som använts behövs för härledd information. Kanske räcker det inte att lägga denna information på typnivå, det kan gälla olika regler för varje förekomst beroende på varifrån den hämtas, under vilka omständigheter, m m.
- Vilket format har en viss datatyp? Eftersom det är fråga om data från många källor kan det vara representerat på många olika sätt: text, grafik, bild, ljud, animation. De kan även vara lagrade på olika medier som papper, disc och tape.
- Vilka tillämpningar använder vilka data? Vilka uppdaterar? För vilket ändamål?

Varje datatyp skulle behöva en beskrivning som bl a gäller:

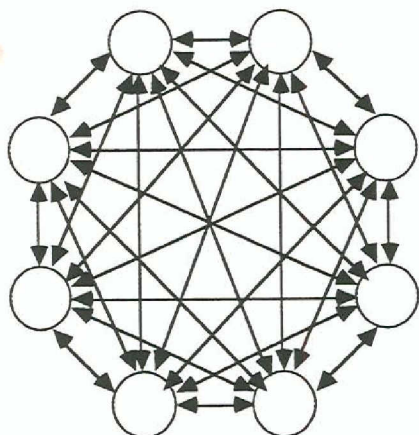
- uppdatering
- ansvar
- aktualitet
- ändringsfrekvens
- version
- kvalitet

Ambitionsnivå får avgöra om dessa uppgifter ska formuleras endast för datatyperna eller även för varje dataelement.

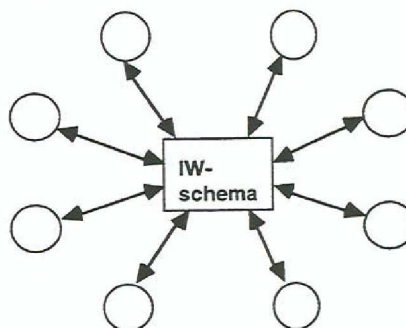
Metadata ligger förmodligen i en s k resurskatalog (repository). Den är uppbyggd på en ambitionsnivå som svarar mot behovet av information hos den berörda verksamheten och informationsarbetaren.

Informationsarbetaren måste kunna arbeta både mot resurskatalogen och mot datamassan i sig. Olika typer av stödjande verktyg behövs i båda fallen för att underlätta arbetet. Det måste finnas ett väldefinierat språk för utsökningar mot datamassan. Språket ska vara anpassat till den begreppsapparat som den verksamhetsövergripande datastrukturen är uppbyggd med (för verktygens operationer och för spontana frågor). Ett motsvarande språk för operationer mot resurskatalogens data är också ett önskemål.

Utan en verksamhetsövergripande datamodell skulle vi i princip behöva definiera en transformation från varje existerande databasschema till var och en av de övriga (se varje pilriktning i figur 1a som exempel). Med en gemensam nämnare begränsas behovet av transformation från IW-schemat till varje databasschema (se figur 1b).



Figur 1a



Figur 1b

Notera att vi med databas menar data som är hanterat av ett visst dbms under ett visst databasschema. Varje databas kan alltså i sig vara distribuerad utan att den ger upphov till mer än en cirkel i bilderna.

Att etablera en verksamhetstäckande datamodell är att etablera en gemensam referenspunkt, en gemensam beskrivning av de begrepp som används och deras samband. Efter det kan man förvänta sig en successiv ensning av begreppsfloran. På så vis kan modellen komma att användas för transformation och därmed kommunikation av data mellan de olika källmiljöerna även vid sidan av gemensamhetsaspekten.

3. Lösningen: Information Warehouse

3.1 Principer

Information Warehouse är IBMs beteckning på en ansats att lösa behoven enligt ovan. Inom engelskspråkig litteratur finns flera olika synonyma beteckningar på det IBM kallar Information Warehouse Database, t ex informational database, informational data store, enterprise level information store. Begreppet information warehouse har blivit populärt (och ofta använt som en generell beteckning för denna typ av teknologi) i och med att IBM i september 1991 annonserade sitt Information Warehouse Framework. Vissa använder istället begreppet data warehouse (något man numer alltid bör göra, eftersom Information Warehouse blivit ett trademark för IBM). Ibland har det legat en nyansskillnad i innebörd, ibland inte. Data warehouse brukar stå för generella åtkomstmekanismer till fristående databaser medan en betoning på information warehouse därutöver antyder en förädling och anpassning av rådata till informationsarbetarens behov. Begreppet warehouse är något missvisande eftersom det lätt kan uppfattas stå endast för en lagringsplats i någon form. Rollen för ett warehouse är i detta sammanhang betydligt vidare, eller som det uttrycks i en IBM-manual (7):

"A set of database management systems, interfaces, processes, tools, and facilities to manage and deliver complete, timely, accurate, and understandable business information to authorized individuals for effective decision making."

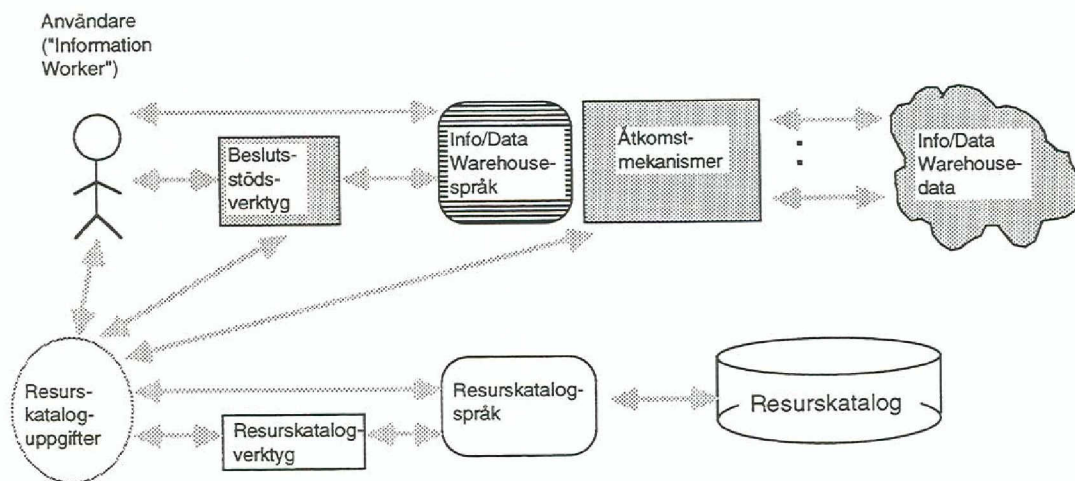
Förutom data, ingår alltså i ett warehouse både mekanismer för att komma åt data och funktioner som ger beslutsstöd.

I fortsättningen kallar vi dessa data för IW-data och den bakomliggande generella datamodellen för IW-schema.

Observera, att tyngdpunkten framförallt ligger på sökningar bland IW-data och dess bakomliggande rådata, inte uppdateringar. I ett längre perspektiv och med en högre ambitionsnivå kan uppdateringar mycket väl bli ett behov. T ex skulle det kunna vara av intresse att lagra fattade beslut/åtgärder/konsekvenser, inklusive referens till de IW-data som legat som underlag.

Ofta är det underförstått, åtminstone i visionen, att IW-data står för verksamhetsövergripande beslutsstödsdata. Men givetvis kan det finnas mellanformer med mer begränsad ambition. Det kan gälla sammanföring av data för en avdelning, ett visst affärsområde, historiska data inom ett visst tidsintervall, data med olika grad av förädling o s v. I samtliga fall kan IW-idén användas. Skillnaden ligger i att dessa mer begränsade IW-data i sin tur kan komma att befinna sig i rollen av rådata för ett mer sammanställt perspektiv.

Sammanfattningsvis är en vital funktion för ett Information Warehouse att kunna tolka frågor i gränssnittets språk, ta reda på var hela eller delar av svaret finns, omformulera delfrågor i enlighet med språkgränssnittet för respektive plats, sända frågor, ta emot svar, förädla och sammanställa dessa till ett totalsvar samt leverera detta till frågeställaren (se figur 2).



Figur 2

Användaren arbetar antingen direkt eller via något verktyg mot IW-data. Både användare, beslutsstöd och åtkomstmekanismerna behöver då och då uppgifter från resurskatalogen för att kunna utföra sina uppgifter. Dessa uppgifter hämtas antingen direkt med hjälp av tillgängligt gränssnittsspråk för resurskatalogen eller via något verktyg. Observera att jag i figuren indikerat resurskatalogen som en enda databas (central eller distribuerad) men rent teoretiskt finns inget hinder för att använda ett IW-resonemang även på dessa typer av data.

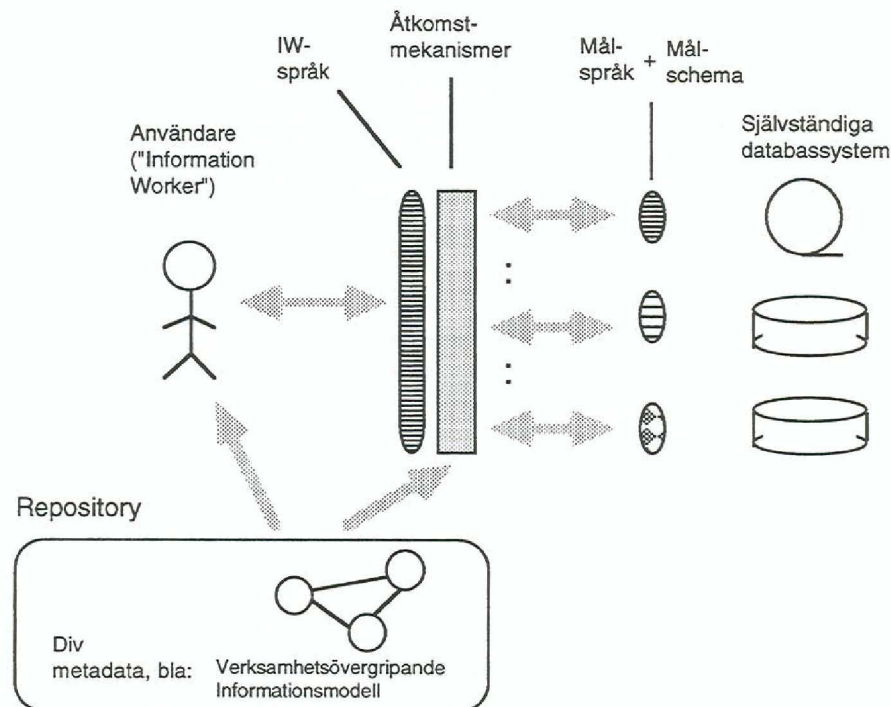
Beslutsstödet lämnar vi därhän i denna rapport eftersom det kan utformas på många olika sätt och efter aktuella behov. Här kan både kommersiella produkter och egenutvecklade tillämpningar komma till användning.

När det gäller att komma åt IW-data kan två olika alternativ användas. Dessa två kan för övrigt kombineras. Antingen ordnar man sökningar i operativa databaser (produktionsdata) på de platser där de finns, eller så ordnar man en separat verksamhetsgemensam IW-databas med de data som antas vara av intresse.

Båda fallen förutsätter att data kan beskrivas på ett semantiskt homogent sätt för användaren utan hänsyn till var de är placerade.

3.2 Sökningar i operativa databaser

Data söks här vid källan, d v s i de operativa databaserna. Sökningen görs i konkurrens med de tillämpningar som används.



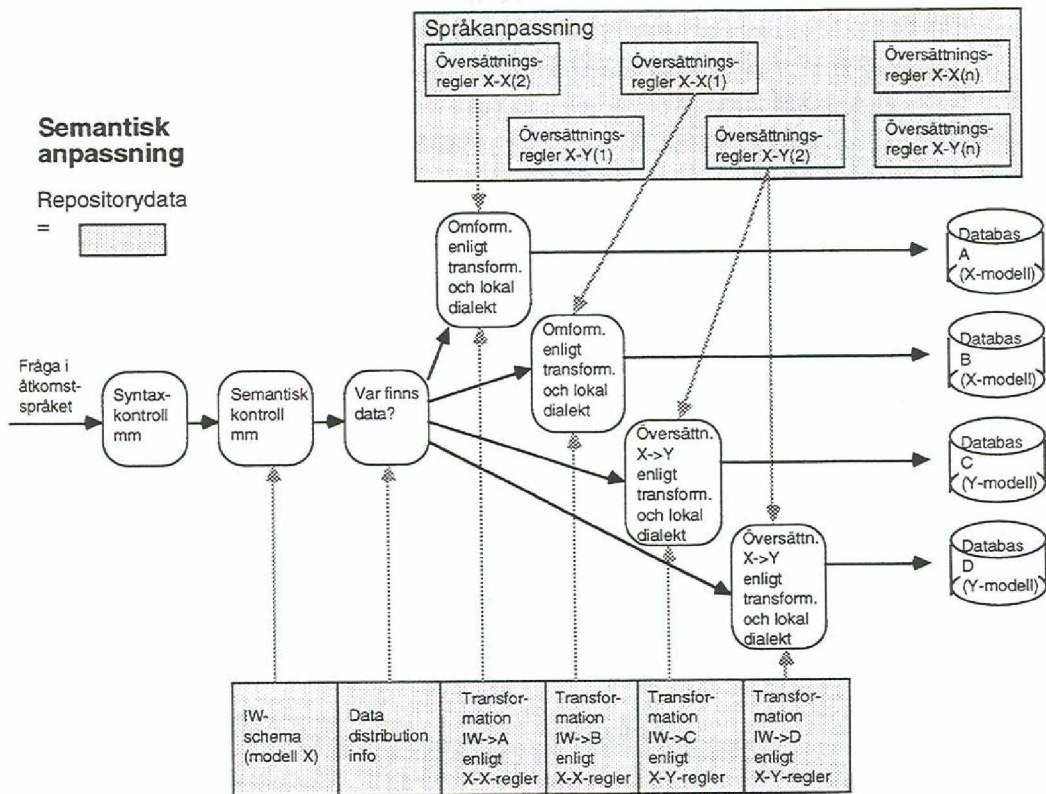
Figur 3

Förutsättningarna för denna typ av sökning är betydligt mer problematiska än de verkar vid ett första påseende. Inte nog med att modelleringsbegreppen kan vara olika (RM, ER, m fl), också med samma begrepp kan frågespråken skilja sig åt eller förekomma i olika dialekter.

Även om inte denna skillnad gäller, måste man förutsätta att ett IW-schema och ett motsvarande källschema skiljer sig åt av olika skäl (olika begrepp, olika upplösning, olika uppfattningar, luckor i semantiken, m m).

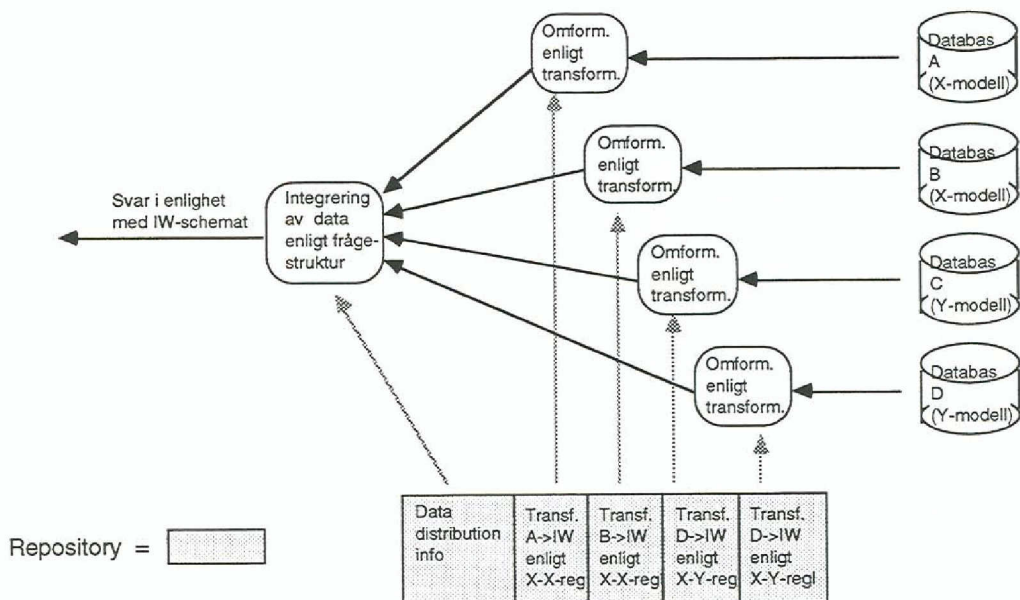
Översättning och/eller transformation kommer nästan undantagslöst att behövas. Endast den som är kunnig i såväl IW-schema som källschema och deras modelleringsbegrepp kan översätta en fråga med IW-begrepp till källbegrepp och transformera ett svar från källdata till IW-data. Därtill kommer behovet av kunskap om de använda frågespråken för att kunna definiera en korrekt översättningsmekanism.

Figur 4 visar ett exempel med fyra källdatabaser där två är uttryckta i X-modelleringsbegrepp och två i Y-modelleringsbegrepp. Språken har olika dialekter (t ex ANSI/SQL, SQL2, ...). Vi utgår från att IW- och källschemana skiljer sig åt både vad gäller semantik och namngivning.



Figur 4

Svaren levereras enligt figur 5.



Figur 5

3.3 Operationer på kopierade data

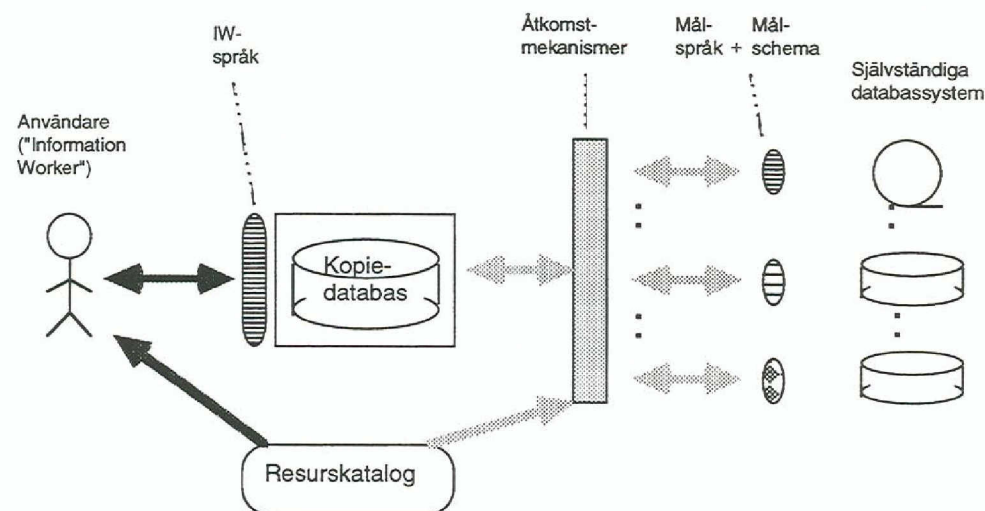
I detta alternativ kopieras intressanta data över från operativa databaser till en separat kopiedatabas, en IW-databas. Kopiedatabasen uppdateras i lämpliga intervaller eller när källdata förändras. Arbetet bör, så långt det är möjligt, göras automatiskt. Kopiedatabasen kan givetvis vara distribuerad, om så önskas. IW-data ligger förberedda, förädlade i enlighet med informationsarbetarens behov.

Kopiering och uppdatering innebär först ett val av de data som bedöms vara av verksamhetsövergripande intresse. Därefter kan rådata eller källdata förädlas för att motsvara de specifika IW-informationsbehoven. IW-databasens volym kan på så vis begränsas. Förädling kan bl a innebära:

- Ett byte av koder mot klartext.
- En översättning av namn- eller formatdialekter till enhetliga standarder.
- Summeringar eller motsvarande funktioner där många ursprungliga dataelement blir ett enda element.
- Grupperingar av data för att tydliggöra valörer, semantik, samhörighet m m.

Kopiedatabasen kan även tillåtas innehålla olika former av historiska data (inga eller endast utvalda raderingar) efter behov, utan att detta alls berör operativa data.

Egentligen är det ingen skillnad mellan de två alternativen om man förutsätter att användaren i båda fallen får jobba mot samma IW-schema. Antingen sker allt arbete när frågan ställs eller vid kopieringen av data. Avser frågorna högggradigt förädlad information kan koplealternativet visa sig vara det enda realistiska alternativet av svarstidsskäl (se figur 6).



Figur 6

3.4 Synpunkter på operativa data

- IW-operationer stör den operativa bearbetningen (pga läsning, accesser, CPU, m m). De kan orsaka prestandaproblem speciellt för tidskritiska tillämpningar.
- IW-operationer kan behöva tillgång till det mesta i en källdatabas. Dessutom finns normalt inslag av spontana frågor och motsvarande spontana behov av tillgång på data. Behörighetskravet kan komma i konflikt med normala behörighetsregler. Någon ökad risk för inkonsistenser borde dock inte finnas så länge IW-operationerna endast gäller utsökningar.
- Detta alternativ innebär ökad sårbarhet vid tekniska störningar. Går en källdatabas ned, kan kanske inte en IW-operation utföras, alternativt ge betydligt lägre kvalitet på svaret. En kopiedatabas är antingen uppe eller ej. Vid störning finns i det fallet, om det anses nödvändigt, alltid möjligheten att ladda en ny kopiedatabas eller utgå från backup.
- * Innan en fråga ställs mot heterogena operativa databaser kan det ibland vara lämpligt att göra en kostnadsuppskattning för att avgöra om svaret är värt kostnaden. Hur gör man denna uppskattning i en heterogen miljö?
- Fördelen är att man, i teorin, alltid får svar baserade på fräscha data. I praktiken ska nog inte detta överbetonas eftersom svaret ändå mycket väl kan vara en förädling av data från flera icke-korrelerade databaser.

3.5 Synpunkter på kopierade data

- IW-data är ofta förädlade från källdata, som i sig kanske inte är av intresse så snart förädlingsarbetet ägt rum. IW-databasens volym kan på så vis begränsas.
- Förädlade data ligger förberedda för presentation. I sökalternativet måste arbetet utföras varje gång dessa data efterfrågas ett slöseri med resurser. Å andra sidan riskerar man i kopiealternativet spendera kraft åt att överföra och förädla data (välja ut, kontrollera, förfina, transportera, packa upp, stoppa in), som aldrig kommer att efterfrågas (åtminstone inte före nästa uppdatering).
- Frågor mot en kopiedatabas kan lättare optimeras.
- En kopiedatabas kan aldrig erbjuda samma aktualitet, eftersom man opererar på gradvis alltmer inaktuella data. Dock, i vissa fall kanske rimlig inaktualitet inte är något problem, speciellt inte i samband med beslutsunderlag av mer övergripande natur.
- Kopiealternativet kan förenkla aspekten dataansvar/ägarskap eftersom olika intressen (operativa - IW) inte står i direkt konflikt med varandra.
- Behörighet kan, eventuellt, specificeras fristående från operativa data.
- Dataadministrationsarbetet renodlas mot kopiedatabasen och överföringen dit.

- Kopiealternativet kräver extra utrymme för sekundärminne.
- En kopiedatabas kan utnyttjas för uppdateringar, när så önskas. Historiska data kan enklare hanteras.

3.6 Bedömning

En grov tolkning av resonemangen i litteraturlistans dokument visar att alternativet med en kopiedatabas för närvarande tycks vara den intressantaste lösningen. En hel del problem av teknisk natur, prestanda m m bortfaller. Informationen koncentreras i en homogen miljö, som lättare kan hanteras och kontrolleras. Risken att IW-databasen kan innehålla något inaktuella uppgifter bedöms inte som allvarlig för de informationsbehov som ska tillgodoses.

4. Frågeställningar att beakta

För att belysa problemkomplexet kring Information Warehouse följer här en, på intet sätt komplett, lista på frågeställningar att ta ställning till och fundera över, inför ett eventuellt beslut om installation av Information Warehouse-konceptet i en verksamhet. Frågeställningarna har grovt grupperats under några olika teman.

4.1 Semantik

Vem är i realiteten kapabel och beredd att definiera begrepp och semantik i ett IW-schema? Även om användarna av IW-data är en mindre grupp människor, som vet vad de vill (m a o kan definiera schemats innehåll) återstår det svåra arbetet att tolka alla källdatabasers scheman och deras korrelation till IW-schemat.

Ett par exempel:

Ordersumma kan betyda många olika saker i olika databaser, bl a:

- Belopp exklusive moms
- Belopp inklusive moms
- Belopp för ej momspliktig kund
- Två belopp: det ena exklusive, det andra inklusive moms
- Belopp för de artiklar som inte är restnoterade
- Nettobelopp efter avdragen kundrabatt
- Belopp enligt kundens uppfattning
- Belopp enligt säljarens dokument

De flesta av dessa varianter kan lösas genom en mer precis namngivning. Inom respektive databas kan man dock inte förvänta sig att finna dessa precisa benämningar. I sitt sammanhang (databas) har *ordersumma* en given semantik. Innan olika databasers ordersummor kan börja blandas i en kopiedatabas, alternativt en fråga ställas mot ett antal källdatabaser, måste ett övergripande tolkningsarbete vara genomfört.

Även med en entydig tolkning kan vissa restriktioner för begreppets användning finnas. Ta *ordersumma* igen:

- Används endast vid order omfattande minst två orderrader.
- Används endast i samband med försäljning av varor per telefon.
- Används endast för order som ännu inte levererats i sin helhet.

I andra sammanhang kan begreppet i sig ha en entydig tolkning men värdet ha en oklar tidsaspekt eller bedömning av när en tillståndsförändring äger rum.

Antal_i_lager för en artikel kan betyda:

- Antal som just nu finns på sin plats i lagret.
- Enligt ovan men minus de som finns på kundorder.
- Enligt ovan men plus de som beställts för inleverans inom en vecka.
- Enligt ovan men minus en chablon för normalt svinn.
- Enligt ovan men plus 25% för att få hem en kundorder i alla lägen.
- Leveransförseningar, klagomål m m fixas senare.

4.2 Ansvar

- Vem ansvarar och administrerar data i en kopiedatabas? De källdataansvariga eller en helt separat kopicansvarig? Om det senare gäller, vad innebär ansvar och hur är det uppdelat mellan källansvarig och kopicansvarig.
- Vem skapar och uppdaterar IW-schemat? Vem ansvarar, kontrollerar och reviderar funktionerna som skapar och uppdaterar motsvarande data? Hur bedrivs arbetet, hur sker kontakten med användarna? Det bör vara någon verksamhetskunnig person eller grupp.
- Hur överförs och revideras behörighetsinformation från källan till den sammanställda databasen? Behörighet till sammanställda data kan eventuellt härledas från behörighet till dess källdata men kan ibland vara frikopplat. Detta behöver nog analyseras för varje situation.
- Vem bedömer nackdelen av att arbeta med kopiedatabasens delvis inaktuella data? Hur sker bedömningen?
- När och hur görs uppdateringar av en kopiedatabas? Vad är lämpligt intervall per källdatabas för att önskad kvalitet ska kunna upprätthållas? Behöver uppdateringarna synkroniseras mellan källdatabaserna? I så fall hur?
- IW-idén kan för vissa upplevas vara ett hot mot den egna kompetensen och yrkesrollen. Beslut kan komma att fattas enligt nya informationsvägar, dvs genom direkt konsultation med IW-databasen. Ibland är det säkert en konstruktiv rationalisering. Ibland kanske man går miste om viktiga bedömningar som bara en erfaren yrkesman kan utföra, i tron på att all kunskap finns i IW. En indirekt följd kan bli en obenägenhet hos yrkesmannen att förse IW-databasen med den information och kunskap han eller hon har.

4.3 Funktion och strategi

- Beslut grundar sig ibland på trender och jämförelser under en tidsperiod. Beslutsunderlag måste i dessa lägen innefatta historiska data. Källdatabaserna innehåller ofta endast aktuella data. IW har på så vis andra behov att tillgodose. Hur kan de tillgodose behovet av historiska data? Hur och när sker uppdateringar av historia? Över vilka tidsperioder? Hur påverkas IW-schemat?
- Samexistens mellan sök- och kopicalternativen kan i vissa fall bedömas vara att föredra. Är det möjligt? Kanske, men det innebär säkert nya problem. Ska t ex ett gemensamt IW-schema användas? Hur hanteras eventuella duplikat?

- Omfattande IW-databaser kan ge upphov till prestandaproblem. Vissa data begärs ofta, andra mycket sällan. Hur optimerar man lagring för bästa möjliga utsökning? Distributionsaspekten? Hur håller man statistik om detta? Kan ändringar i den fysiska lagringen göras dynamiskt med hänsyn till denna statistik? Vilka kriterier?
- Ofta behövs IW-data till mycket avgörande beslut. Kvaliteten är avgörande. Hur kan man garantera denna kvalitet, speciellt om den baseras på data från många håll? Hur kan man övertyga användaren om den? Vem tar ansvar? Om kvaliteten är tveksam kan det vara säkrare och bättre att inte grunda beslut på IW-data.
- Källdatabaserna är som levande väsen. De ändras inte bara till sitt innehåll utan även till sin uppbyggnad och semantik. Hur får IW reda på förändringar? Hur meddelas dessa? Vem känner ansvar? Vilka arbetssteg bör genomföras vid en sådan anpassning?
- Källdatabaserna lever inte isolerade liv. Det är mycket sannolikt att de kom munerar data mellan sig. Samma uppgifter kan alltså finnas i flera databaser, kanske under olika namn och med olika aktualitet. Hur tar IW hänsyn till denna omständighet?
- Hur hanterar man en situation där data finns i databas a och borde finnas i databas b, men inte gör det? Ska frågan inte alls besvaras, besvaras med varning, eller ...?
- En IW-strategi kan riskera att konservera gammal, diversifierad teknik och datahantering och därmed motverka förnyelseprocesser mot mer samordnade lösningar.
- Resurskatalogdelen är ett problem för sig. Hur ska dess innehåll struktureras? Vilka gränssnitt ska gälla? Informationsarbetarna måste t ex enkelt kunna finna uppgifter om vilka IW-data som finns tillgängliga. Att navigera i en grafisk representation av datastrukturen kan räcka för vissa, andra kanske vill ha strukturen uppdelad per affärsområde och enligt typiska informationsbehov, andra återigen vill söka efter ansvarsområden, osv. Ingen standard finns vare sig beträffande modelleringsbegrepp, gränssnitt eller innehåll.
- Samma resonemang som föregående kan användas för IW-databasen. Antagligen kommer det att finnas flera konkurrerande IW-ansatser på marknaden, var och en med sin produktuppsättning. Även om varje verksamhet ursprungligen endast etablerar ett IW, förändras behov och verksamheter (slås ihop/delas upp) över tiden. Följaktligen uppstår så småningom behov av kommunikation och samordning mellan olika IW-ansatser. Tillbaka till "gå" igen?
- Behovet av IW baseras på övergripande resonemang kring information som en resurs i verksamheten, behovet av heltäckande kunskap, behovet av fullödiga beslutsunderlag. Vissa anser att endast sammanställd information är av intresse. Andra menar att man börjar sin analys på en sammanställd nivå, fokuserar sitt intresse på något specifikt, vill ha mer detaljerad information om detta, hittar något specifikt på denna nivå osv, kanske ända ner till källdata.
- Finns det mer detaljerade analyser genomförda som avser vilka personalkategorier och/eller funktioner som har vilka informationsbehov från en IW-databas? Hur vill de kunna jobba?

- Hur övertygar man i konkreta termer budgetansvariga om lönsamheten i en IW-satsning? Argumenten tenderar bli abstrakta (bättre kvalitet, bättre beslut, snabbare uppföljning).
- Passar IW i den existerande informationssystemstrategin för verksamheten? Kanske är alternativet en helt ny, mer integrerad plattformstrategi? Kanske bör existerande databaser få leva sitt liv och nya databaser och tillämpningar successivt inlemmas i en välgenomtänkt IW-strategi?

4.4 Frågegränssnitt

Varje typ av IW utnyttjar sina modelleringsbegrepp och sitt gränssnittsspråk vid formulering av IW-schema och IW-frågor. IBM har valt att basera sin IW-ansats på SQL och relationsmodellen. I anslutning till detta uppstår bl a följande frågor:

- Att IBM valt SQL innebär inte automatiskt att marknaden som helhet går den vägen. Objektorienterade synsätt vinner alltmer terräng. Även IBM har, bl a i resurskatalogsammanhang, valt Entity-Relationship-modeller och objektmodeller. Standarder är en förutsättning för att marknaden ska acceptera IW-idén. Ännu finns ingen klar trend.
- IW ska bl a användas som beslutsstöd. Enkla, kraftfulla gränssnitt är ett måste. SQL är inte lämpligt som slutanvändargränssnitt. SQLs fördel är att det är standardiserat, känt och allmänt accepterat. Sannolikt kom mer vi att se mer av grafiska gränssnitt baserade på någon form av objektorienterad modell för den spontana frågeverksamheten. Hur länge dröjer det tills ett nytt språk uppnår samma stabilitet som SQL? Kanske är den spontana frågeverksamheten överbetonad. För dem som ändå vill arbeta med ett speciellt verktyg är det inte av intresse vilken gränssnittsstandard vergtyget använder mot IW. Det är snarare en angelägenhet för dem som realiserar verktyg och IW!
- Det finns idag ett antal verktyg som inte arbetar med SQL och data enligt relationsmodellen. Ska ett sådant verktyg i IW-miljö behöva gå igenom följande procedur?
 - Omformulera sina frågeställningar till SQL,
 - anropa ett IW-verktyg som
 - översätter från SQL vidare tillbaka till databasens gränssnitt,
 - tar emot svaret i icke-relationsform,
 - översätter till relationsform,
 - överlämnar resultatet till det ursprungliga verktyget som
 - översätter resultatet enligt eget behov (d v s enligt databasens gränssnitt),
 - och till sist kan börja jobba?
- Ca 80-90% av all världens data anses fortfarande ligga i egenutvecklade, hierarkiska eller nätverksdatabaser. Frågan är om SQL i detta perspektiv är ett så nödvändigt val med hänsyn till marknaden kunskap om frågespråk. Är det alldeles givet att SQL är så lämpligt för formulering av frågor mot icke-relationsdata? Det kanske är lika bra att ta en helt ny ansats, som ligger mer i linje med behovet av en gemensam informationsmodell!?

5. Slutsatser

Än så länge saknas praktisk erfarenhet av IW som är satt i system. Konceptet är mycket komplext och kommer därför att behöva utvecklas under lång tid. Man ska akta sig för en övertro på teknikens förmåga att lösa bekymren. Problemen ligger snarare inom tolkning och semantik, inom strategianpassning, kunskapsuppbyggnad och attityder. Kanske måste hela datahanteringen i en verksamhet ses över innan en fungerande realisering blir möjlig.

En ansats som är grundad på en kopiedatabas och parvisa överenskommelser mellan IW-instansen och en källdata-leverantör kan vara en rimligt ambitiös interimslösning. Överenskommelsen regleras i ett kontrakt som fastställer villkor, struktur, semantik och innehåll för de data som ska överföras. Som överföringssyntax kan någon marknadsetablerad syntax användas, t ex EDIFACT. Var och hur leverantören hämtar sina data för leverans är dennes ensak, så länge som villkoren i kontraktet uppfylls. På marknaden finns även verktyg för detta ändamål. Se bl a referenser i bilagan.

När IBM nu har annonserat sitt Information Warehouse måste man hålla i minnet att det är en partsinlaga – ett ramverk som successivt kom mer att fyllas på med produkter. IBMs inresse inom området får den normala sidoeffekten att området hamnar mer i fokus, blir av mer direkt intresse för såväl konkurrenter som leverantörer av kompletterande produkter. Kunskaper och erfarenheter kommer att skapas snabbare än vad som annars skulle vara fallet.

En annan viktig allmän sidoeffekt blir förhoppningsvis att information som sådan får högre status som en vital resurs i en verksamhet. Man kommer att fokusera mer på behoven hos informationsarbetare. Detta i sin tur leder, återigen förhoppningsvis, till att företagen mer noga går igenom och definierar de datatyper som existerar och som behövs i verksamheten. Därmed läggs även grunden för mer semantiskt genomtänkta och för verksamheten stödjande databaser. Kanske blir följderna på sikt mer enhetlig semantik mellan fristående databaser och därmed enklare, säkrare transformationer.

För övrigt, IW är egentligen inget nytt fenomen. Inom samma problemkomplex men under beteckningen "integration of heterogeneous databases" pågår både forskning, utveckling och debatt. Utopia-projektet på Posten är värt en eloge i dessa sammanhang.

Ett viktigt första steg är att lära mer om IW och dess syften, möjligheter och fallgropar. Ett andra steg är att försöka ta fram en verksamhetsgemensam, övergripande informationsmodell (IW-schema). Lyckas man där (vilket inte alls är säkert) kan det vara idé att börja tänka på en teknisk lösning i flera steg. Börja med en begränsad mängd data och en begränsad mängd användare i ett pilotprojekt. En lugn och välgenomtänkt expansion ska hela tiden baseras på återkoppling av kunskap och erfarenheter.

Behoven tycks uppenbara, teknik och verktyg börjar komma, strategier för införande och användning är oklara, vägen från vision till verklighet tycks lång, men på något sätt ändå gripbar. Eller som det så fyndigt uttrycks i en IBM-broschyr:

*Before your next decision,
here's one thing you should know.
Everything.*

Litteraturförteckning

1. *Information Warehouse - An Introduction*, IBM manual GC26-4876-01, 1992.
2. *SAA - Introduction to Distributed Data*, IBM manual GC26-4831-00, 1991.
3. Delvin, Murphy; *An architecture for a business and information system*, IBM Systems Journal, vol 27, no 1, 1988.
4. Sayles; *Universal Dataspeak Becomes a Relational Reality*, Database Programming & Design, august 1992.
5. Goldring; *Delivering Data to the Information Warehouse*, InfoDB, summer 1992.
6. *IBM's Information Warehouse - Just Another Framework?*, Database review, oktober 1991.
7. *EDA/SQL in an Information Warehouse: Early installation experiences*, IBM manual GG24-3779, may 1992.
8. *Information Warehouse - Implementation Overview*, IBM manual GG24-3869, december 1992.
9. Inmon; *EIS and the Data Warehouse*, Database Programming & Design, november 1992.
10. Bozman; *DB2 to get client books*, Computerworld, no 50, 1992.
11. *DataHub: window on the Information Warehouse*, Okänd publikation.
12. *Information Warehouse Architecture*, IBM manual.

Bilaga:

IBMs Information Warehouse Framework

Innehåll

B1. Översikt 20

- B1.1 Verksamhetsdata 20
- B1.2 Tillämpnings- och beslutsstödssystem 21
- B1.3 Dataleverans 21
- B1.4 Sammanställning 22

B2. Distribuering enligt IBM 23

B3. DRDA-alternativet 24

B4. EDA/SQL 24

B5. Till sist 26

Detta är en genomgång av IBMs Warehouse Framework. Syftet är att ge en mycket översiktlig introduktion av dagsläget vid årsskiftet 92/93. Såväl idéer och produkter är under kontinuerlig utveckling.

Brasklapp: Innehållet i denna bilaga är grundat på tillgänglig och efter bästa förstånd läst litteratur, se litteraturlistan. Missuppfattningar kan förekomma. Den som behöver mer exakt och helt aktuell information bör ta direktkontakt med IBM.

B1. Översikt

IBMs Information Warehouse Framework annonserades i september 1991. Vid den tidpunkten var det en idéskiss, ett ramverk med det långsiktiga syftet att tillgodose ett växande behov av att se och arbeta med information som en verksamhetsövergripande resurs. Annonseringen refererade även till några produkter, vars egenskaper gör dem till naturliga komponenter i ramverket. Dit hör:

- databashanterarna DB2/MVS, SQL/DS, OS/400, DB2/2 och DB2/6000
- DRDA (Distributed Relational Database Access) som är en client-server mekanism för relationsdata
- EDA/SQL som är en produkt från samarbetspartnern Information Builders Inc
- datakopieringsprodukterna Data Propagator (DProp), Data Extract Facility (DXT).

Målsättningen formuleras i följande citat ur en IBM-manual [1]:

"The IBM Warehouse Framework defines a set of database management systems, interfaces, tools, and facilities that manage and deliver reliable, timely, accurate, and understandable business information to authorized individuals for business decision making."

En anpassning till befintliga och framväxande industristandarder eftersträvas. Erfarenhetsbaserad utveckling ska ske i samverkan med International Alliance Members.

IBM delar upp ramverket i tre komponenter:

- *Enterprise Data* (inkl deras aktuella dbms, bl a DB2, SQL/DS, OS/400, OS/2)
- *Data Delivery Mechanisms* (access, copy, transport, transformation, enrichment, et c)
- *Applications and Decision Support Systems for Information Workers* (Notera begreppet "information workers".)

B1.1 Verksamhetsdata

I filosofin ligger att verksamhetsdata (enterprise data) ska kunna vara av vad slag som helst och ligga lagrad under i stort sett vilket dbms eller filsystem som helst. Bekymren kring tolkningen av datas semantik inom och mellan databaser har ännu inte behandlats fullt ut.

Relationsmodellen förutsätts vara det använda modelleringsspråket.

Att ta fram en övergripande verksamhetsmodell kräver ett långsiktigt, metodiskt arbete med fokus på informationsarbetarens informationsbehov. För att utröna om behoven kan tillgodoses, kan t ex existerande relationsdatabasers scheman vara en informations-

källa. Olika reverse engineering-produkter finns för att ta fram datastrukturer ur befintliga system, strukturerade på annat sätt. I en IBM-manual [8] nämns t ex reverse engineering-produkter från Bachman och Knowledgeware. Man refererar även till existerande modeller framtagna för olika branscher, t ex banker och försäkringsbolag.

B1.2 Tillämpnings- och beslutsstödssystem

Tillämpnings- och beslutsstödssystem (applications and decision support systems) kan vara allt från specifika tillämpningar till verktyg som finns tillgängliga på marknaden. Verktygen kan vara inriktade mot stöd vid arbete på datamassan i sig. IBM-manualen [8] exemplifierar med bl a AS (Application System), Lotus 1-2-3, QMF (MVS/Query Management Facility), QM (OS/2 Query Manager) och DIS (Data Interpretation System).

Användaren behöver också stöd som syftar till att öka kunskap om data och om var den finns, dvs arbete mot en resurskatalog (repository) av beskrivningsinformation (s k metadata). Inte minst inom den senare kategorin behövs verktyg för att IW-visionen ska kunna bli en realitet för informationsarbetare. Till denna kategori hör BIL (Business Information Locator) [10]. Vi kan förmoda att flera verktyg med samma syfte så småningom kommer ut på marknaden. IW, AD/Cycle och Systemview har ju här delvis överlappande behov.

Generellt kan sägas att stödjande verktyg är speciellt viktigt för dem, som i sin övriga verksamhet inte handskas med aktuell typ av information och sådana som inte är vana vid att se och tolka datamodeller direkt. För andra kanske SQL eller någon grafiskt baserad E-R-modell är fullt tillräckligt, åtminstone för spontan frågeverksamhet.

B1.3 Dataleverans

Data i IW struktureras i enlighet med relationsmodellen (SQL). Det betyder dock inte att beslutstödssystem måste tillämpa SQL i gränssnittet mot användaren. Användargränssnittet blir det som respektive verktyg tycker är lämpligt. Däremot måste verktyget uttrycka sig i termer av SQL gentemot de funktioner som har att leverera IW-data. Med andra ord, SQL används för att modellera IW-data och som utsökningsspråk. Det verksamhetsövergripande IW-schemat måste av den anledningen vara uttryckt i termer av relationsmodellen.

Som diskuterades i första delen, kan behovet av tillgång på data tillgodoses på två principiellt olika sätt. Antingen sker sökning i produktionsdata eller i en kopiedatabas.

Alternativet med produktionsdata tillgodoses för närvarande enligt följande. Den person som befinner sig i en IBM-miljö under SAA och önskar operera på relationsdata med hjälp av SAA SQL, använder sig lämpligen av Distributed Relational Database Architecture (DRDA). Av prestandaskäl bör man alltid använda DRDA när det är möjligt. I övriga fall används Enterprise Data Access/SQL (EDA/SQL). Se utförligare beskrivning av DRDA och EDA/SQL nedan.

Laddning och uppdatering av en kopiedatabas kan givetvis utföras genom någon tillämpning som man skrivit själv, vilken indirekt kanske utnyttjar DRDA eller EDA/SQL. Det finns även produkter som på begäran eller med automatik kan utföra motsvarande uppgifter.

Bland dessa återfinns:

Data Extract Facility (DXT) flyttar data i någon form in i en relationsdatabas. DXT kopierar när så begärs eller med givna intervaller. DXT kan även använda EDA/SQL vilket öppnar kontaktytan mot många olika typer av produkter som inte är utvecklade av IBM.

Data Propagator (DProp) överför IMS-data till DB2-data i samband med uppdatering av IMS-databasen. Den ger med andra ord en synkronisering mellan original och kopia i form av "delta"-uppdateringar.

Remote Recovery Data Facility (RRDF) används primärt för recovery av externa DB2-data men kan även användas för uppdatering av en DB2-baserad kopiedatabas.

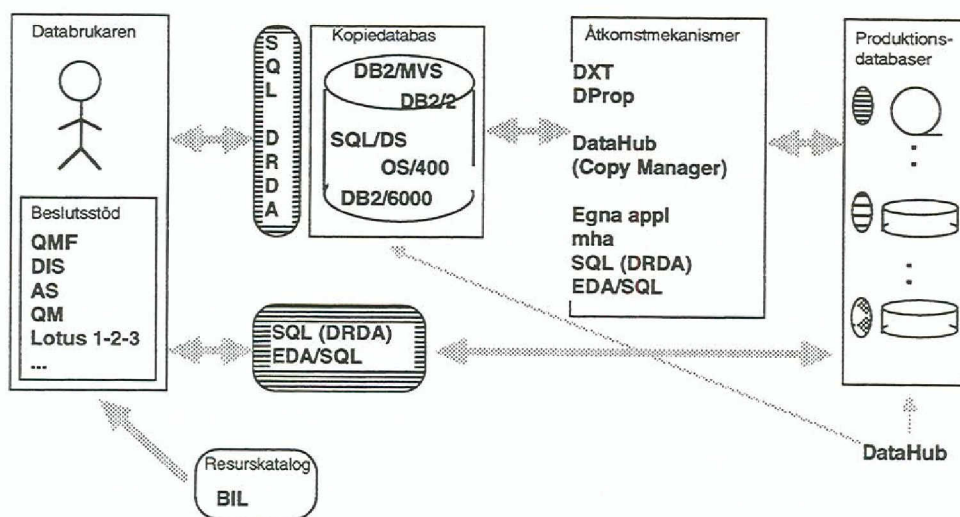
Åtkomst till en laddad IW kopiedatabas sker som tidigare nämnts lämpligen med hjälp av SQL och vid behov DRDA.

Inom detta område finns även *DataHub* [11], en produkt för administration av databaser. Den kan bl a användas för att:

- Flytta data mellan olika databaser oavsett om de ligger under DB2, SQL/DS, OS/400 eller OS/2.
- Hantera behörigheter.
- Göra backup och omorganisera databaser.

B1.4 Sammanställning

Figur b1 sammanfattar de olika produkterna och deras placering i ramverket.



Figur b1

B2. Distribering enligt IBM

Några ord om IBMs syn på distribuerade databaser hämtade från en manual [2].

Definitioner

unit of work (UOW) "Any number of SQL statements followed by either a single COMMIT or ROLLBACK SQL statement".

database environment A database management system and at least one database.

database A schema and corresponding data.

schema The definitions of all active base tables, viewed tables, indexes, privileges, and user names.

data Every value in every active table in the schema.

Fyra ambitionsnivåer

Remote request består av en enda SQL-sats som i sin tur avser data från ett och samma ställe (database environment). Man måste veta adressen till denna.

Remote units of work (RUW) alla SQL-satser i en UOW måste referera till data lagrade på ett och samma ställe. Det motsvarar en tillämpning som körs på samma maskin som det lokala systemet för databashantering. För att utföra denna sammansatta operation behövs på klient-sidan en requester som sköter kontakterna med det aktuella fjärrdatabashanteringssystemet. I princip kan en tillämpning tolka data som lokala genom att en requester utför det nödvändiga externa arbetet. Distributed Relational Database Architecture (DRDA) har RUW-egenskaper över heterogena miljöer.

Distributed unit of work (DUW) Inom ett och samma UOW kan referenser ske till data (tabeller), som finns på flera fysiska platser (olika database environments). Restriktionen är att varje SQL-sats måste referera data inom ett environment. Går inte det måste satsen delas upp i flera. Fysisk adress till DBMS måste finnas. DB2 Version 2 RELEASE 3 klarar DUW men med restriktionen att en UOW bara får innehålla uppdateringar mot ett environment.

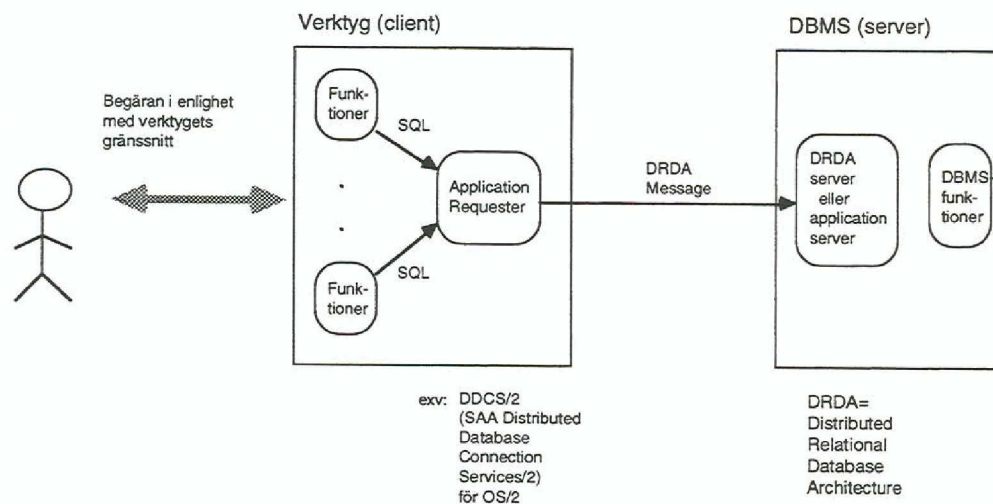
Distributed request (DR) ållåter en viss SQL-sats att referera data under olika dbms. Man kan operera som om allt låg lokalt utan tanke på den reella distributionen.

För alla alternativen har gällt att varje tabell uppträder som en separat, komplett tabell inom sitt environment. Ibland kan verkligheten vara mer komplex än så, eller åtminstone önskar man att man kunde skapa mer komplexa situationer. T ex:

- *Extracts* är kopior av hela eller delar av tabeller. Begärda av användare.
- *Snapshots* är kopior av tabeller som endast kan läsas och som skapas och uppdateras av systemet.
- *Replication* är snarlika snapshots men med möjlighet till uppdatering. Uppdateringar skickas dessutom till de övriga databaser som innehåller kopierade data enligt önskemål. Bra t ex om man vid vissa noder vill hålla en uppdaterad lokal kopia av någon intressant tabell.
- *Distributed table* är en fysisk splittring av en tabell horisontellt och/eller vertikalt.

B3. DRDA-alternativet

DRDA-nivån är fullt ut specificerad och som sådan en del av IBMs IW Framework. Figur b2 beskriver bearbetningsgången.



Figur b2

Varje program eller verktyg med en application requester (DRDA Requester), som kan kommunicera med en application server enligt DRDA-protokoll, fungerar under Framework. Samma sak gäller på server-sidan. En requester tar emot en SQL-sats och förpackar den i ett DRDA Message för vidare transport till en DRDA Server där satsen packas upp och köras.

Klient-sidan måste, förutom schemat, känna till den dialekt av SQL som används på server-sidan, eftersom ingen översättning äger rum. Eventuella behov av förädling av data innan leveransen till användaren ombesörjes genom verktygsfunktioner.

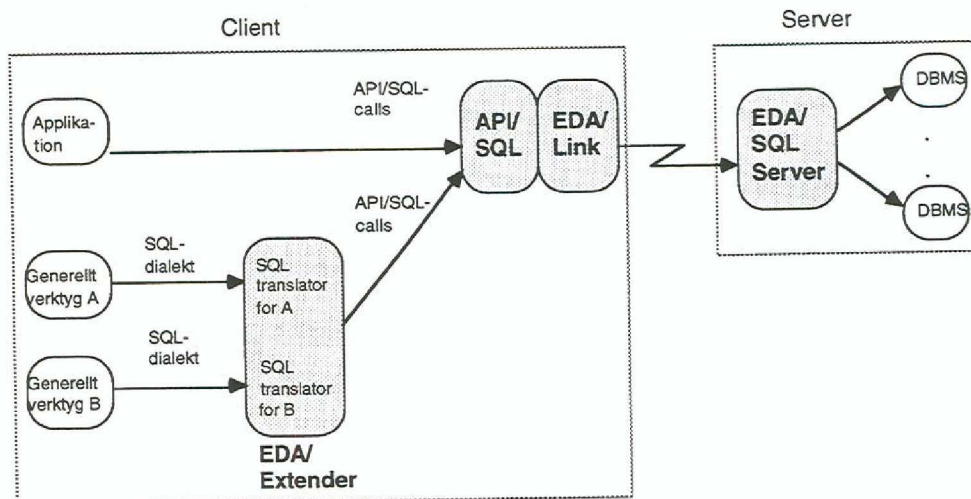
För den som inte vill skriva en egen DRDA Requester tillhandahåller IBM Distributed Database Connection Services/2 (DDCS/2). På serversidan har följande IBM-produkter DRDA Server Support: DB2, SQL/DS, OS/400. Dessa kan också fungera som DRDA Requester. Flera fristående dbms-leverantörer har eller kommer att erbjuda DRDA Server Support.

B4. EDA/SQL

EDA/SQL är en produkt utvecklad av Information Builders Inc (IBI). IBI tillhör IBMs International Alliance.

EDA/SQL klarar av att söka data oavsett om de är strukturerade i enlighet med relationsmodellen eller ej och oavsett vilket gränssnittspråk aktuella databaser använder. För närvarande klarar man enligt uppgift ett 50-tal olika databashanterare (bla TOTAL, ORACLE, IDMS, ADABAS, DB2, IMS, SYBASE).

EDA/SQL består av fyra delar: API/SQL, EDA/SQL Server, EDA/Link, EDA/Extender sammankopplade enligt figur b3.



Figur b3

API/SQL

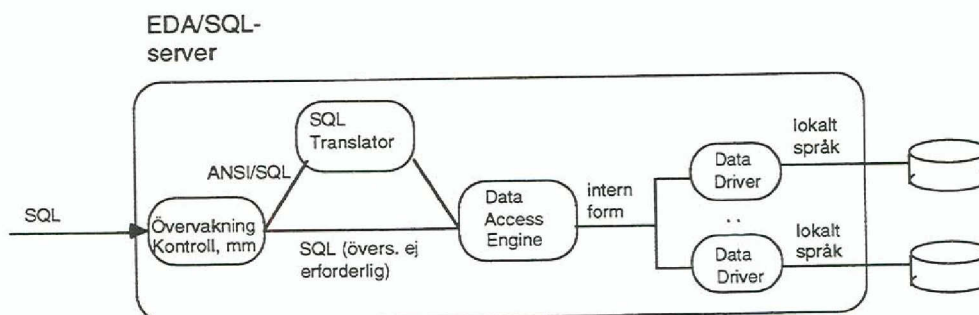
API/SQL är ett antal funktioner som tolkar instruktionsrepertoaren för EDA/SQL-gränssnittet. Instruktionerna är administrativa, kommunikations- och databasorienterade. Bland de senare finns instruktioner för att begära körning av SQL-satser och även körning av procedurer som är lagrade i servern (s k remote procedure calls) samt för att ombesörja hantering av resultat som ska tillbaka till frågeställaren. SQL-satsen återfinns i textform som en parameter i instruktionen (ett s k call-level interface).

Bland administrativa instruktioner finns initiering och avslutning av tillämpningars kontakt med EDA/SQL.

Upprättande av kontakt med aktuella servrar återfinns bland instruktionerna för kommunikation.

EDA/SQL Server

En EDA/SQL Server kan ha många klienter samtidigt uppkopplade mot sig. Förutom att administrera och prioritera inkommande förfrågningar från klienterna är den centrala arbetsuppgiften att se till att inkommen SQL-sats utförs på korrekt sätt. Figur b4 visar vad som händer.



Figur b4

SQL-satsen är normalt formulerad i enlighet med ANSI/SQL-standard (dvs oavsett var data finns och hur data är strukturerat). Refererade data förutsätts namngivna i enlighet med måldatabasernas terminologi. SQL-satsens refererade data matchas mot internt lagrad kunskap om vilka data som finns var (från EDA/SQLs egen interna resurskatalog) samt översätts till en intern form.

Den interna formen skickas vidare till Data Access Engine som ser till att alla berörda måldatabaser efterfrågas på sina delsvår samt sammanställer (joins m m) dessa till ett totalsvar.

Varje typ av DBMS har sin Data Driver som anpassar internformen till exakt de krav respektive DBMS-gränssnitt ställer.

Den som vill formulera sin fråga i enlighet med den version av SQL, som gäller för måldatabasernas DBMS, kan göra detta och samtidigt begära att översättningssteget hoppas över (för effektivare körning).

EDA/Link

EDA/Link fungerar som kommunikationsgränssnitt mellan klient och server.

EDA/Extender

Dataåtkomst är inte bara av intresse för tillämpningar man skrivit själv. Många existerande verktyg (t ex spreadsheet, CASE) och tillämpningar arbetar mot olika typer av databaser. EDA/Extender består av en modul per anpassat verktyg/databashanterare. Uppgiften är att anpassa olika existerande verktygs och tillämpningars normala sätt att operera mot databaser till API/SQL-anrop när så önskas. På så vis öppnas automatiskt vägen till databaser som annars kanske inte skulle kunna nås, åtminstone inte utan ett omfattande realiseringsarbete inom respektive produkt.

Antag en SQL-tillämpning. I samband med ett SQL-anrop går EDA/Extender in och kontrollerar om tabellreferenserna avser tabeller i en EDA/Server (alternativt att en tidigare 'CONNECT-sats' öppnat en aktuell EDA/Server). Om inte, utförs operationerna mot den lokala databasen som vanligt. I annat fall packas frågan i enlighet med API/SQLs instruktionsrepertoar varefter API/SQL öppnar förbindelse med aktuell EDA/Server med hjälp av EDA/Link. Svaret levereras motsatta vägen. Extenders finns för flera miljöer. Flera är på gång.

B5. Till sist

I en IBM-manual [7] diskuteras på ett mycket informativt och utförligt sätt hur införandet av IW lämpligen kan läggas upp.

Produkter inom Information Warehouse utvecklas inte inom någon viss grupp vid IBM. Ramverket ska snarast ses som en idéskiss för ett viktigt insatsområde. Vissa produkter kan naturligt inordnas under ramverket, andra ligger vid sidan om.

Idéutvecklingen sker vid Santa Teresa-laboratoriet, San José. Lämpliga produkter hämtas från många håll inom IBM-koncernen liksom från samarbetspartners. Bakom IBMs IW-satsning finns med andra ord en ansenlig kapacitet och, får vi hoppas, en långsiktig målsättning.

TIDIGARE UTGIVNA PUBLIKATIONER AV TRIADGRUPPEN

Verksamhetskrav på informationsadministration

- V 1: IA och verksamhetskrav – erfarenheter från offentlig förvaltning
- V 2: Fallstudie av IA-projektet vid Televerket
- V 3: IA-erfarenheter från företag och myndigheter

Modellering

- N 1: Modelleringsansatser för begrepps- och datamodellering: – Beskrivning och försök till jämförelse
- N 2: Generering av konceptuella modeller från policydokument
- N 3: Espritprojektet Tempora
- N 4: Prövning av regelbaserad metodik inom Posten
- N 5: En kokbok i remodellering - utkast
- N 6: Datorstöd för modellintegration
- N 7: Modellbaserad kunskapsinsamling
- N 8: Modellkvalitet
- N 9: Samband mellan dokument och modeller

Utbildning

- H 1: Handledarutbildning för modelleringsledare, avancerad
- H 2: Slutrapport HUMLA prototyp

Uttagssystem

- U 1: Hybris i Unix-miljö
- U 2: DEBRIS

Katalogprinciper

- K 1: IRDS
- K 2: IRDS Modeller och modellnivåer
- K 3: Koppning begreppsmodell - relationsmodell
- K 4: IBM:s Repository Manager - en Introduktion
- K 5: IBM:s Repository Manager: Datamodelleringsbegreppen
- K 6: IBM:s Repository Manager: Begreppsmodellering i Information Model
- K 7: IBM Repository Manager: Attribut- och värdemodellering i Enterprise Submodel
- K 8: Navigering i Repository
- K 9: TRIAD Newsletter – IRDS inom ISO. Dagsläget
- K 10: TRIAD Newsletter –ISO/IRDS. Händelseutvecklingen 91/92
- K 11: Samverkan mellan resurskataloger – visioner eller behov
- K 12: AD/Cycle I Information Model – Processer och informationsflöden mellan processer
- K 13: AD/Cycle I Information Model – Info Flows inom Processmodellen
- K 14: AD/Cycle I Information Model – Relationsdatabasmodellering
- K 15: AD/Cycle I Information Model – Härlednings-specifikationer i begreppsmodellen
- K 16: IA-prototyp
- K 17: Repository AD/Cycle - International Users Group
- K 18: RAD-konferensen i Chicago, 1992
- K 19: Vad händer inom ANSI-IRDS?

KORT OM TRIAD

Triad är namnet på ett treårigt samarbetsprojekt kring informationsadministration och dataadministration, IA/DA, som Televerket, Posten, Ericsson, Statskontoret och SISU bedriver. Syftet är att utveckla parternas synsätt, metoder och hjälpmedel inom detta område.

Arbetet inom Triad är uppdelat i delprojekt som är sammanförda i tre block.

Beställarblocket vänder sig dels till dem som är verksamhetsansvariga och måste ta ställning till IA/DA-satsningar, dels till dem som har ansvaret för IA/DA inom en organisation. Delprojekten inom detta block arbetar med att formulera verksamhetens krav på IA/DA samt studerar och beskriver roller, organisation och arbetsformer för IA/DA-arbete.

Utförarblocket vänder sig till dem som arbetar med IA/DA. Delprojekten arbetar med modellering, data- och resurskataloger samt uttagssystem.

Kunskapsförmedling är det block som ser till att resultaten kommer Triad-parterna till godo. Detta sker bland annat genom kurser, seminarier samt genom att rapporter, som denna, ges ut.
