

## Data Warehousing i et Forsikringselskab: En kronik i Politiken

af Morten J. Lintrup



Morten Johan Lintrup  
mj@topdanmark.dk

*Denne artikel fortæller ikke, hvordan man skal lave data warehousing i et forsikringselskab, og den er slet ikke en kronik i dagbladet Politiken. Men den beskriver formålet og nogle overordnede principper for data warehousing opgaven i et forsikringselskab. – Og så er det jo altid godt med en iøjnefaldende overskrift – også i et fagtidsskrift – selv om så det kun kan blive i underrubrikken. Og berettigelsen ligger i at artiklens form, længde og krav til læseren er søgt holdt analogt til kriterierne for en kronik i Politiken med det formål at formidle et fagligt emne til en motiveret og veluddannet men også bredere orienteret læserkreds.*

### Data Warehousing generelt

Data Warehousing (DW) som begreb opstod i USA i løbet af 1980'erne og er i dag suppleret med den nogenlunde synonyme betegnelse Business Intelligence (BI). Derudover findes en lang række andre forkortelser og betegnelser for opgaver og systemer af mere begrænset rækkevidde og holdbarhed, undertiden udviklet af og knyttet til bestemte softwareproducenter. Disse snævrere og mere specifikke begreber vil ikke blive behandlet her. Ingen nævnt, ingen glemt. Lad det blot være sagt, at sådanne begreber typisk betegner enten en delopgave inden for data warehousing eller et tilgrænsende område, som man af den ene eller anden grund kan have valgt at udskille som en separat funktion i virksomheden. Det kan også forekomme lidt misvisende at tale om "ét" data warehouse, da de fleste større

virksomheder i praksis har flere mindre DW enheder, som dækker hver sine funktioner. Teorien anbefaler ganske vist et enkelt samlet data warehouse, men realiteten er normalt anderledes, og det har mange forfattere og foredragsholdere på området da også taget til efterretning uden nævneværdig fortrydelse.

En større virksomhed – her et forsikringselskab – har typisk en række mere eller mindre selvstændige forretningsenheder, som alle benytter IT-teknologi i betydeligt omfang. IT er i grunden en stabsfunktion, men har alligevel en særstilling i forhold til de andre, oprindelige stabsfunktioner, som regnskab, økonomi, marketing, m.v., fordi IT er blevet så fundamental og integreret en del af hver enkelt medarbejders hverdag. Men hvad

*Morten Johan Lintrup er Civilingeniør, HD og arbejder som Data Warehouse ingeniør i Business Intelligence, Topdanmark Forsikring A/S.*

er så data warehousing, og hvor findes data warehousing funktionen i virksomheden?

### Hvorfor Data Warehousing

Behovet for data warehousing som selvstændigt begreb voksede frem med udviklingen mod stadig større og mere diversificerede virksomheder, kombineret med bevægelse hen imod mere netværksorienterede organisationsformer og understøttet af stadig mere og kraftigere IT, både hardware og software. Der opstod en erkendelse af at regnskabsopgaven grundlæggende bestod i at tilfredsstille omverdenens krav til virksomhedens økonomiske styring, og økonomifunktionen understøttede de tilsvarende interne behov for overblik over virksomhedens pengestrømme, men der manglede alligevel noget...

Der var særlige funktioner til dybtgående og specialiseret analyse på udvalgte områder, fx med aktuarmæssig statistik. Men for overblik og effektiv styring behøvede man også en dækkende indsamling og registrering på tværs af virksomheden af andre oplysninger end de snævert økonomiske, det være sig antalsopgørelser, forbrugsregistrering, produktionsopfølgning eller lagerstyring. I sig selv opgaver, som var kendt fra tidligere; det nye var alene mængden af data og behovet for at skabe overblikket på tværs af forskellige forretningssenheder, – og det med en passende lille forsinkelse til at muliggøre rettidig ledelsesmæssig handling.

For at forstå indholdet af data warehousing, kan man også overveje den mere moderne betegnelse BI, Business Intelligence. Business betyder selvfølgelig 'forretning', men Intelligence har her betydningen 'efterretninger', ligesom i fx navnet på den amerikanske efterretningstjeneste, CIA, Central Intelligence Agency.

### Indhold i Data Warehousing

Det er den forretningsmæssige baggrund for og formål med data warehousing. Lige så snart man dykker ned i det arbejdsmæssige indhold i data warehousing, så viser det sig dog at være meget IT tungt. Deri ligger utvivlsomt forklaringen på at mange virksomheder har overladt opgaven til IT organisationen og dermed lagt den et andet sted, end hvor den reelt hører hjemme. Og det er i sin tur utvivlsomt en del af forklaringen på at relativt mange data warehouse initiativer historisk må siges at være slået fejl. Omvendt må det selvfølgelig medgives, at med opgavens store indhold af IT, så er tilknytningen den vej også meget vigtig. Og det er ikke nogen naturlov, at tingene går galt, hvis data warehousing organisatorisk placeres dér. Det er bare et dårligt udgangspunkt. Heller ingen vil jo normalt placere regnskabs- eller økonomifunktionen som en del af IT. Men i den sidste ende afhænger resultatet selvsagt af de mennesker, som arbejder i data warehouse funktionen, og af samarbejdet med resten af organisationen.

Ifølge den organisatoriske teori for data warehousing er de største risici ved at placere opgaven i IT dels at den forankres for langt fra de ledere i forretningen, som skal bruge resultaterne, – og dels at fokus i selve udførelsen flyttes fra det egentlige, forretningsmæssige formål til selve den tekniske kompleksitet i hverdagen. Rent praktisk kan man også sige, det kan være besnærende at "overlade IT til IT medarbejdere". Men IT og IT kan være flere ting. Ikke mindst et forsikringselskab har en lang række tunge operationelle systemer, som bare skal fungere dag ud og dag ind uden afbrydelser. Det har man lært og vænnet sig til og indrettet sig efter med en række sunde principper og politikker og systemer, som er udvalgt og indkøbt for at honorere det krav. Men for et data warehouse er mange ting en hel del anderledes end i de operationelle systemer. DW systemerne behøver ikke nødven-



Vignetterne viser frimærker sammentrykt med forsikringsreklamer. Danmarks første reklamesammentryk udkom september 1927 med reklame for Hafnia og Danske Phønix. Allerede 1911 var i Tyskland udsendt reklamesammentryk for Limania cykel-ulykkesforsikring.

digvis være til rådighed hele døgnet, hver eneste dag. Lejlighedsvis ”nedetid” og ditto forsinkelse kan være helt acceptabelt. Blandt andet det forhold kan have stor betydning for hvilke arbejdsmetoder og hvilke systemer, det kan være hensigtsmæssigt at benytte.

### En sammenligning

På dette sted kommer jeg til at tænke på min første ansættelse som nyuddannet ingeniør. Det var i Hedeselskabet som hydrolog. Hedeselskabet beskæftigede sig også med skovdrift. Skovene har jo mange funktioner, og der er mange slags mennesker, som færdes i dem. Skovarbejderen arbejder på et bestemt sted på et givet tidspunkt, og det er vigtigt at have tænkt på og sørget for alt muligt sikkerhedsudstyr, for der er store kræfter i spil, og der kan ske alvorlige ulykker, hvis noget går galt. Orienteringsløberen derimod skal netop kunne komme hurtigt rundt over det hele og er tilsvarende afhængig af let udstyr. Snubler hun, så kan hun som regel bare rejse sig op og løbe videre, og træder hun uforvarende i et

vandhul, så er det ubehageligt – men heller ikke mere. På mange måder kan man sammenligne traditionelt IT arbejde med det tunge skovarbejde og data warehousing med orienteringsløbet. En dygtig skovarbejder og en dygtig orienteringsløber ved hver især meget om skoven, og noget af det en skovarbejder har lært, kan han bruge i et orienteringsløb. Men skovarbejdet er også omgærdet med veltjente principper og indgroede vaner, som det er nødvendigt at aflægge i et orienteringsløb.’

### Hardware og softwarwe

Tilbage ved data warehousing opgaven er der eksempelvis betydelig forskel på datamodelering til traditionel, operationel brug og til data warehousing. Traditionelt bruges såkaldt entitets-relations modellering (ER), men til data warehouse formål bruges primært såkaldt dimensionel modellering. Hvor ER modeller kan siges at have til formål at minimere og ideelt helt eliminere redundans i data, så er kontrolleret redundans en nødvendig og

integreret del af dimensionel modellering.

Tilsvarende med hensyn til hardware og software. Stabilitet og pålidelighed er alfa og omega for operationel IT. Stabilitet og pålidelighed er selvfølgelig også vigtige parametre i data warehousing, men det er hastighed og tilgængelighed i mindst lige så høj grad. Med tilgængelighed menes her omtrent det samme som det lidt forslidte begreb brugervenlighed. Et system, brugerne ikke kan forstå, forekommer dem utilgængeligt. Det vil derfor ikke blive brugt. Den afgørende forskel er at operationelle systemer typisk benyttes intensivt og næsten dagligt af brugerne, så de lærer eventuelle besynderligheder og accepterer dem. Brugen af data warehouse systemer kan ofte være mere ekstensiv og meget gerne også mere eksplorativ end operationel.

Et konkret eksempel, hvor hensigtsmæssige principper for operationel IT og data warehousing kan komme i konflikt, kunne være brugen af "MS-Ware". "Microsoft bashing" er en hæderkronet beskæftigelse for mange inkarnerede IT folk. Nægtes kan det jo heller ikke, at pålidelighed og sikkerhed med mange Microsoft produkter har ladet ganske meget tilbage at ønske, når de blev frigivet. Men modsat det indtryk mange gerne vil give, så har Microsoft aldrig blot ladet stå til, men har løbende bestræbt sig – og er lykkedes med – at udvikle og forbedre sine produkter væsentligt.

Principiel modstand mod MS produkter – måske lige bortset fra Office pakken – bidrager til at cementere skellet mellem IT brugere i forretningen og IT administratorer i IT organisationen. Det er ikke nødvendigvis hensigtsmæssigt. Til data warehousing kan det tværtimod være hensigtsmæssigt at kvalificerede brugere i forretningsorganisationen får adgang til avancerede funktioner og software, der er designet så det ligner kendte systemer fra hverdagen og ligger i logisk forlængelse af disse. Mange uafhængige software-leverandører bestræber sig også på at udvikle syste-

mer med næsten samme brugeroplevelse som Windows og MS Office, så Microsoft er ikke den eneste spiller i sin ende af banen.

### **Data Warehousing i praksis**

---

Indtil nu har artiklen kun handlet om organisation og overordnede principper. Afslutningsvis skal vi derfor se på et eksempel helt nede i detail- og hverdagsniveau, som illustrerer hvorfor den organisatoriske teori for data warehouse er som beskrevet.

Data afleveres eller hentes ind fra mange forskellige systemer og forretningsenheder til virksomhedens data warehouse. Der ligger typisk en stor opgave i at tjekke og kvalitets-sikre disse data, men den opgave skal her blot nævnes i forbifarten; – den kan eventuelt blive genstand for en senere, selvstændig artikel. – En vis del af data vil være beskrivende i form af en lang række forretningskoder med tilknyttede tekster og forklaringer. En væsentlig opgave er at harmonisere, det vil sige forlige, beskrivende data med forskellige formater, koder og tekster fra flere systemer, når disse beskrivende data reelt betyder det samme.

Denne opgave kan principielt tænkes løst på forskellige måder. Ved en "ren" IT orienteret og informationsteoretisk tankegang er det nærliggende at mene, det kun handler om loyalt at gengive beskrivende tekster, som de fødes i kildesystemerne. "Intet må trækkes fra eller lægges til; informationsindholdet kan ikke øges, kun reduceres". Men dette synspunkt forflygtiger fuldstændig, at data warehouse faktisk har en forretningsmæssig opgave og ansvar for at skabe overblik. Og rå data i sig selv giver netop ikke basis for overblik og indsigt.

Næste trin i en IT orienteret tankegang vil være, at "alle fejl skal rettes, hvor de opstår, så det er altid de ansvarlige for kildesystemerne, der skal ændre i deres koder og beskrivelser". Og det er jo i hvert fald ikke helt forkert. Men desværre er det heller ikke helt rigtigt. I prak-

sis vil grænsen nemlig være flydende mellem fejl, mangler og mere eller mindre nødvendige forskelligheder. Objektive fejl skal selvfølgelig principielt rettes i kildesystemerne. Men hvis den organisatoriske træghed kan betyde, at det kan tage et halvt år at få rettet en stavfejl, så ville det være forkert at medvirke til at udbrede fejlen gennem data warehouse så længe. Et andet eksempel kan være gamle tekster fra en mainframe tid, hvor alt blev skrevet med store bogstaver. Er det en fejl eller en mangel i dag? – Tja, ... det bedste er jo nok bare at se at få det rettet frem for at diskutere det som et problem. I nogle tilfælde kan det også være forbundet med reelle vanskeligheder at rette fejl ved kilden. En kort beskrivende tekst kan måske være forkert i data warehouse, men nødvendig at bibeholde i et gammelt kildesystem, fx fordi systemet fungerer med en snæver begrænsning på længden af tekststreng, som var gældende, da systemet i sin tid blev udviklet.

Vi kan gå videre til spørgsmål om beskrivende tekster, som opstår i forbindelse med harmonisering af data på tværs af kildesystemer. Et sted benyttes måske teksten ”Forsikringstager flyttet” – et andet sted kan det hedde ”Kunden flyttet”. I lige netop sådan et tilfælde kan det være nemt at henholde sig til en generel sprogpolitik, der måske fastsætter at det første er forkert og det andet er rigtigt. Men i mange andre tilfælde kan der fx være tale om at det samme produkt hedder noget forskelligt på forskellige markeder, eller når det sælges gennem forskellige afsætningskanaler. Den slags forskelle er i sig selv udtryk for en bevidst og nødvendig strategi, når et stort selskab skal kunne tilpasse sig og agere efter en lokal markedssituation. Den slags forskelle er derfor på ingen måde udtryk for fejl i kildesystemerne og kan ikke bare „rettes” ved kilden.

## Konklusion

Det korte af det lange ved diskussionen om beskrivende koder og tekster er at data warehouse funktionen skal være bevidst om at forholde sig til meget mere end det IT-mæssige indhold i at indsamle, gemme og flytte data. I et data warehouse skal man være indstillet på og parat til at tage ansvar også for forretningsmæssige spørgsmål i tilknytning til den fysiske indsamling og formidling af data fra mange steder i organisationen. Det er blandt andet derfor den organisatoriske teori for data warehousing med en vis stædighed fastholder, at et data warehouse primært hører hjemme i selve forretningen og ikke i IT organisationen.

Artiklens indhold er alene forfatterens ansvar og en afspejling hans læsning og erfaringer.

### Udvalgt relevant litteratur

- English, Larry P: Improving Data Warehouse and Business Information Quality – Methods for Reducing Costs and Increasing Profits, Wiley publ. 1999, 544 p.
- Gonzales, Michael L: IBM Data Warehousing – With IBM Business Intelligence Tools, Wiley publ. 2003, 704 p.
- Kimball, Ralph og Margy Ross: The Data Warehouse Toolkit, 2<sup>nd</sup> Ed., Wiley publ. 2002, 440 p. [www.kimballgroup.com](http://www.kimballgroup.com)
- Moody, Daniel L and Mark A R Kortink: From ER Models to Dimensional Models, Part I: Bridging the Gap between OLTP and OLAP Design. Business Intelligence Journal V 8 (3) pp. 7-24, 2003.
- Moody, Daniel L and Mark A R Kortink: From ER Models to Dimensional Models, Part II: Advanced Design Issues. Business Intelligence Journal V 8 (4) pp. 20-29, 2003.
- Thomann, J: Total Quality Management for Data Warehousing, 2003. Kursusmateriale for TDWI, The Data Warehousing Institute. [www.tdwi.org](http://www.tdwi.org)
- Todman, Chris: Designing A Data Warehouse, Supporting CRM, Hewlett-Packard Professional Books 2001, 329 p.