Datavarehus håndbok

[Ralph Kimball](http://www.bedreinnsikt.no/innhold/ralph-kimball) introduserte i 1996 boken [The Data Warehouse Toolkit](http://www.bedreinnsikt.no/innhold/verkt%C3%B8ykasse-design-av-datavarehus) med teknikker for dimensjonal modellering av datavarehus og har senere forfinet dette til en beste praksis. Her vil du finne svar på det meste som du vil komme til å lure på under designfasen av et datavarehus.

Innhold

[Identifisering av forretningsprosessene 3](#_Toc11680876)

[Bussarkitektur for bedriftens datavarehus 4](#_Toc11680877)

[Dimensjonell modelering av datavarehus 4](#_Toc11680878)

[Grunnleggende konsepter 4](#_Toc11680879)

[Samle forretningsmessige krav og data realiteter 4](#_Toc11680880)

[Dimensjonal modellering workshops med kunden 4](#_Toc11680881)

[Dimensjonal design prosess i 4 steg 5](#_Toc11680882)

[Forretningsprosesser 5](#_Toc11680883)

[Detaljeringsgrad 5](#_Toc11680884)

[Dimensjoner som definerer kontekst 6](#_Toc11680885)

[Fakta fra målinger 6](#_Toc11680886)

[Stjerne skjemaer og OLAP kuber 6](#_Toc11680887)

[Utvidelser til dimensjonal modeller 7](#_Toc11680888)

[Grunnleggende teknikker for fakatabeller 7](#_Toc11680889)

[Strukturen til en faktatabell 7](#_Toc11680890)

[Additive, Semi-additive, og ikke-additive fakta 8](#_Toc11680891)

[NULL verdier i faktatabeller 8](#_Toc11680892)

[Ulike fakta for samme måling 8](#_Toc11680893)

[Transaksjonsfaktatabeller 8](#_Toc11680894)

[Periodiske faktatabeller 9](#_Toc11680895)

[Akkumulerende øyeblikksbilde faktatabeller 9](#_Toc11680896)

[Faktaløse faktatabeller 9](#_Toc11680897)

[Aggregerte faktatabeller eller kuber 10](#_Toc11680898)

[Konsolidert faktatabeller 10](#_Toc11680899)

[Grunnleggende teknikker for dimensjonstabeller 10](#_Toc11680900)

[Strukturen til en dimensjonstabell 10](#_Toc11680901)

[Surrogatnøkler i dimensjonstabeller 10](#_Toc11680902)

[Naturlige, holdbare, og supernatural nøkler 11](#_Toc11680903)

[Drille ned 11](#_Toc11680904)

[Degenererte dimensjoner 11](#_Toc11680905)

[Degenererte flate dimensjoner 12](#_Toc11680906)

[Flere hierarkier i dimensjoner 12](#_Toc11680907)

[Flagg og indikatorer som dimensjonelle tekstattributter 12](#_Toc11680908)

[Null-attributter i dimensjoner 12](#_Toc11680909)

[Tidsdimensjonen 12](#_Toc11680910)

[Dimensjon for roller 15](#_Toc11680911)

[Oppsamlingsdimensjoner 15](#_Toc11680912)

[Snøfnuggdimensjon 15](#_Toc11680913)

[Støttedimensjoner 16](#_Toc11680914)

[Integrasjon via overenstemte dimensjoner 16](#_Toc11680915)

[Samsvarende dimensjoner 16](#_Toc11680916)

[Reduserte opprullingsdimensjoner 16](#_Toc11680917)

[Drille på tvers 17](#_Toc11680918)

[Verdikjede 17](#_Toc11680919)

[Bussarkitektur for datavarehus 17](#_Toc11680920)

[Bussmatriser for datavarehus 17](#_Toc11680921)

[Mulighet/interessent matrise 18](#_Toc11680922)

[Teknikker for sakteendrede dimensioner 18](#_Toc11680923)

[Type 0: Beholde originalen 18](#_Toc11680924)

[Type 1: Overskriv 18](#_Toc11680925)

[Type 2: Legg til ny rad 18](#_Toc11680926)

[Teknikker for dimensjonshierarki 19](#_Toc11680927)

[Posisjonelle hierarki med fast dybde 19](#_Toc11680928)

[Hierarkier med litt variabel dybde 19](#_Toc11680929)

[Hierarkier med variabel dybde 19](#_Toc11680930)

[Arkitektur for ETL-prosessen 20](#_Toc11680931)

[Oversettelse av engelske begreper 20](#_Toc11680932)

# Identifisering av forretningsprosessene

I Kimballs tilnærming er det fire viktige beslutninger når du utformer en dimensjonal modell for datavarehuset; identifiser forretningsprosessen, detaljeringsgraden, dimensjoner og fakta. Selv om dette høres trivielt ut så viser det seg at team ofte snuble på det første trinnet. De sliter med å artikulere forretningsprosessen som det er et begrep som synes å ha ulik betydning avhengig av konteksten. Siden definisjonen av forretningsprosessen er den første staven i bakken når du utformer en dimensjonal modell ønsker vi å eliminere forvirring i vår kontekst.

La oss begynne med å diskutere hva en forretningsprosess er og ikke er. Når du utformer en dimensjonal modell så refererer ikke forretningsprosessen til en virksomhets avdeling, organisasjon eller funksjon. Likeledes bør den ikke refererer til en enkelt rapport eller en bestemt analyse.

For utvikleren av en dimensjonal modell er forretningsprosessen et hendelse eller en aktivitet som genererer eller samler målinger. Disse verdiene er resultatmålene for organisasjonen. Forretningsanalytikere vil uunngåelig ønsker å granske og vurdere disse målingene av tilsynelatende ubegrensende kombinasjoner av filtre og begrensninger. Som dimensjonale modellbyggere er det vår jobb å presentere disse målingene i en lettforståelig struktur som responderer raskt på uforutsigbare forespørsler.

Ved identifisering av forretningsprosessen for dimensjonal modelleringdukker det ofte opp noen felles kjennetegn og mønstre.

Forretningsprosesser er vanligvis støttet av et operativt system. For eksempel støttes forretningsprosess for fakturering støttes av et faktureringssystem; likeledes for innkjøp, bestilling og forretningsprosesser for mottak.

Forretningsprosesser generere eller samle inn unike målinger med et unik detaljnivå og dimensionalitet som brukes til å måle sine resultater. Noen ganger er målingene et direkte resultat fra forretningsprosessen. Andre ganger er målingene avledninger. Uansett leverer forretningsprosessene resultattall som brukes av ulike analytiske prosesser. For eksempel støtter forretningsprosessen for salgsbestilling en rekke rapporter og analyser som for eksempel kundeanalyse, salgsrepresentant ytelse og så videre.

Forretningsprosesser blir ofte uttrykt som handlingsverb med tilhørende dimensjoner som substantiver som beskriver hvem, hva , hvor, når , hvorfor og hvordan relatert til prosessen . For eksempel vil det fakturaforretningsprosessenresulterer være skiver - og - terninger og analysert etter dato, kunde , service / -produktet, og så videre.  
Forretningsprosesser er vanligvis utløst av en inngang og resulterer i produksjonen som må overvåkes . For eksempel , er en akseptert forslaget innspill til bestillingsprosessen som resulterer i en salgsordre og tilhørende beregninger . I dette scenariet , er forretningsprosessen salgs bestilling ; vil du ha en ordre faktum bord med salgsordren som en potensiell degenerert dimensjon og ordrebeløp og teller som fakta . Prøv å forestille seg den generelle flyten fra innspill til en forretningsprosess , som resulterer i utgangs beregninger . I de fleste organisasjoner er det en rekke forretningsprosesser hvor utganger fra en prosess blir input til den neste . I parlance av en tredimensjonal modellbygger , vil disse prosesser resultere i en serie av faktatabeller .  
Analytikere noen ganger ønsker å bore tvers av forretningsprosesser , ser på resultatet av en prosess sammen med resultatene av en annen . Drilling gjennom prosesser er absolutt levedyktig hvis de dimensjoner som er felles for begge prosessene er dannet .  
Bestemme organisasjonens sentrale forretningsprosesser er avgjørende for å etablere det samlede rammen av dimensjonale modeller . Den enkleste måten å finne disse prosessene er ved å lytte til forretningsbrukere. Hvilke prosesser generere resultattall de er mest interessert i overvåking ? Samtidig bør datavarehus -teamet skal vurdere realitetene i kildekode miljøet for å levere data ettertraktet av virksomheten .

En siste kommentar ... Det bør gå uten å si at den stadig populære dashbordet er IKKE en forretningsprosess ; Den presenterer resultatene av en rekke individuelle forretningsprosesser.

[Design Tip #69 Identifying Business Processes](http://www.kimballgroup.com/2005/07/05/design-tip-69-identifying-business-processes/)

# Bussarkitektur for bedriftens datavarehus

The Kimball konsernets Enterprise Data Warehouse Bus Arkitektur er et sentralt element i vår tilnærming. Introdusert i 1990-årene, gjør at teknologien-og database-uavhengig buss arkitektur for inkrementell datavarehus og business intelligence (DW / BI) utvikling. Det dekomponerer DW / BI planprosessen i håndterbare biter ved å fokusere på organisasjonens sentrale forretningsprosesser, sammen med de tilknyttede dannet dimensjoner.

# Dimensjonell modelering av datavarehus

## Grunnleggende konsepter

### Samle forretningsmessige krav og data realiteter

Før du starter med dimensjonal modellering, må teamet forstå forretningsbehovene og realitetene i de underliggende kildedata. Du avdekke kravene via sesjoner med representanter fra foretaket for å forstå deres mål basert på nøkkelindikatorer, forretningsmessige problemstillinger, beslutningsprosesser og behov for støtte i analyseprosessen. Samtidig blir realiteter i dataene avdekket med møter med utviklerne med kjennskap til kildesystemene og høynivå dataprofilering for å vurdere hva som er realistisk å gjennomføre med hensyn på de tilgjengelige dataene.

### Dimensjonal modellering workshops med kunden

Dimensjonale modeller bør utformes i samarbeid med fageksperter fra forretningen og personer med oversikt og ansvar for dataene. Datamodellereren er ansvarlig, men modellen skal utfolde seg via en rekke svært interaktive workshops med representanter fra virksomheten. Disse workshops tilbyr en mulighet til å detaljere kravene med virksomheten. Dimensjonale modeller bør ikke være utformet isolert av folk som ikke fullt ut forstår virksomheten og deres behov; samarbeid er avgjørende!

### Dimensjonal design prosess i 4 steg

Disse fire beslutningene er det viktig at man må bestemme seg for under utformingen av en dimensjonal modell for datavarehuset:

1. Velg forretningsprosessen
2. Definer detaljeringsgraden
3. Identifiser dimensjonene
4. Identifiser faktaene

Disse spørsmålene besvares ved å vurdere behovene fra virksomheten sammen med realitetene i de underliggende datakildene, som man kommer frem til under de samarbeidende modelleringsøktene. Etter forretningsprosessen og identifisering av dimensjoner/fakta vil teamet bestemme seg for navn på tabeller/kolonner, innhente eksempelverdier og forretningsregler.

Det er viktig at representanter fra kunden, både fra forretningssiden og de med kjennskap til dataene, deltar i denne designaktiviteten slik at alle er enige om designvalgene.

### Forretningsprosesser

Forretningsprosesser er de operasjonelle aktivitetene utført av organisasjonen, slik som å tar imot en bestilling (ordre), behandler et forsikringskrav, registrering av studenter for en klasse, eller avstemning av hver konto hver måned. Hendelser i forretningsprosesser generere eller fanger opp resultattall som oversettes til fakta i en faktatabell. De fleste faktatabeller fokusere på resultatene av en enkelt forretningsprosess. Valg av prosess er viktig fordi den definerer et bestemt designmål som muliggjør definisjoner av grain, dimensjoner og fakta. Hver forretningsprosess tilsvar en rad i bedriften datavarehus bussmatrise.

### Detaljeringsgrad

Deklarasjon av kornet er det avgjørende skrittet i et dimensjonal design. Kornet fastsetter nøyaktig hva en enkelt rad i faktatabell representerer. Fastsettelse av kornet blir en bindende kontrakt på designet. Kornet må erklæres før du velger dimensjoner eller fakta fordi hver kandidater til dimensjoner og fakta må være i samsvar med kornet. Denne konsistensen sikrer et uniformt dimensjonalt design som er kritisk for BI-løsningens ytelse og brukervennlighet. Det atomiske kornet refererer til det laveste nivået av data som fanges opp av en gitt forretningsprosess. Vi oppfordrer deg til å starte med å fokusere på atomiske-spesifiserte data fordi det tåler angrepet av uforutsigbare brukerspørsmål; oppsummerte korn er viktig for ytelse tuning, men de forantar virksomhetens vanlige spørsmål. Hver foreslåtte korn til en faktatabell resulterer i et separat fysisk tabell; forskjellige korn må ikke blandes i samme faktatabell.

### Dimensjoner som definerer kontekst

Dimesjoner gi "hvem, hva, hvor, når, hvorfor, og hvordan" konteksten til hendelser i forretningsprosesser. Dimensjonstabeller inneholde beskrivende attributter som brukes av BI-løsninger til å filtrere og gruppere fakta. Med kornet til en faktatabell fremst i tankene kan alle aktuelle dimensjoner identifiseres. Hvis mulig bør hver rad i en faktatabell kun være assosiert med en enkelt verdi i dimensjonstabellen.

Dimensjonstabeller blir noen ganger kalt "sjelen" til datavarehus fordi de inneholder de inngangspunkter og beskrivende etiketter som gjør at DW/BI systemet kan utnyttes for forretningsanalyse. Uforholdsmessig mye arbeid blir lagt inn i datastyring og utvikling av dimensjonstabeller fordi de er driverne av brukerens BI opplevelse.

### Fakta fra målinger

Fakta er målingene som kommer fra hendelser i forretningsprosessen og er nesten alltid numeriske. En rad i faktatabellen har et en-til-en forhold til et måltall tilknyttet en hendelse definert av at kornet (detaljeringsgraden) til faktatabellen. Dermed vil faktatabellen korrospondere med en fysisk observert hendelse, og ikke ut fra kravene til en bestemt rapport. Kun fakta som er konsistente med det definerte kornet er tillatt i faktatabell. For eksempel, i en salgstransaksjon, antall produkter som selges og dens stykkpris er gode fakta, mens butikksjefen lønn er forbudt.

### Stjerne skjemaer og OLAP kuber

Dimensjonale modeller fokuserer på prosess hendelser målinger, hvor data deles inn i enten fakta eller beskrivende kontekster for "hvem, hva, hvor, når, hvorfor eller hvordan".

Dimensjonale modeller kan implementeres i både relasjonsdatabaser, kalt stjernediagram, eller flerdimensjonale databaser, kjent som online analytical processing (OLAP) kuber. Karakteristikken til et stjerneskjema er at det består av faktatabeller knyttet til tilhørende dimensjonstabeller via primærnøkkel/fremmednøkkel relasjoner. OLAP kuber kan være tilsvarende i innhold til, eller oftere avledet fra, et relasjonsstjerneskjema. En OLAP kube inneholder dimensjonale attributter og fakta, men er tilgjengelig via språk med mer analytiske muligheter enn SQL, for eksempel XMLA. OLAP kuber er inkludert i denne listen av grunnleggende teknikker fordi en kube er ofte det siste trinnet i en dimensjonal DW/BI systemdistribusjon, eller kan eksistere som en samlet struktur basert på et mer atomisk relasjonsstjerneskjema.

Ordet "Kimball" er synonymt med dimensjonal modellering. Ralph Kimball oppfant ikke de opprinnelige grunnleggende begrepene fakta og dimensjoner, men han etablerte en omfattende portefølje av dimensjonale teknikker og vokabular, inkludert conformed dimensions, slowly changing dimensions, junk dimensions, mini-dimensions, bridge tables, periodic and accumulating snapshot fact tables, og mange andre begreper.

Dimensjonal modellering passer ikke bare for organisasjoner som omfavner Kimball arkitektur men også de som følger Corporate Information Factory (CIF) arkitektur forfektet av Bill Inmon og andre. Dimensjonal modellering beste praksis er arkitekturnøytral.

For å få et raskt overblikk på dimensjonal modellering kan vi anbefale deg å lese gjennom følgende artikler som er tilgjengelig på nettsidene til Kimball Group:

* [A Dimensional Modeling Manifesto](http://www.kimballgroup.com/1997/08/02/a-dimensional-modeling-manifesto/), DBMS, August 1997
* [Fact Tables and Dimension Tables](http://www.kimballgroup.com/2003/01/01/fact-tables-and-dimension-tables/), Intelligent Enterprise, January, 2003
* [Dividing the World](http://www.kimballgroup.com/2008/02/20/dividing-the-world/), Data Management Review, March 2008
* [Essential Steps for the Integrated Enterprise Data Warehouse, Part 1](http://www.kimballgroup.com/2008/03/17/essential-steps-for-the-integrated-enterprise-data-warehouse-part-1/), Data Management Review, April 2008
* [Essential Steps for the Integrated Enterprise Data Warehouse, Part 2](http://www.kimballgroup.com/2008/04/11/essential-steps-for-the-integrated-enterprise-data-warehouse-part-2/), Data Management Review, May 2008
* [Kimball’s Ten Essential Rules of Dimensional Modeling](http://www.kimballgroup.com/2009/05/29/the-10-essential-rules-of-dimensional-modeling/), Intelligent Enterprise, May 2009

Vi anbefaler også at du skaffer deg en kopi av boken [The Data Warehouse Toolkit](http://www.bedreinnsikt.no/innhold/verkt%C3%B8ykasse-design-av-datavarehus) av Ralph Kimball som gir deg en komplett beskrivelse av dimensjonal modellering med utallige eksempler fra den virkelige verden.

### Utvidelser til dimensjonal modeller

Dimensjonale modeller er tilpasningsdyktige når data og datarelasjonene endres. Alle de følgende endringene kan gjennomføres uten å endre noen av de eksisterende BI spørringene eller applikasjonen, og uten noen av søkeresultatene endrer seg:

* Fakta som er konsistente med kornet til en eksisterende faktatabell kan legges til ved å opprette nye kolonner.
* Dimensjoner kan legges til en eksisterende faktatabellen ved å opprette nye kolonner for fremmednøkler, så lenge de ikke endre kornet til faktatabellen.
* Attributter kan legges til en eksisterende dimensjonstabellen ved å opprette nye kolonner.
* Kornet til en faktatabell kan gjøres mer atomisk ved å legge til attributter i en eksisterende dimensjonstabell, og deretter omarbeide faktatabellen til et lavere korn, samtidig som man passer på bevare navnet på de eksisternede kolonnene i faktatabellen og de tilhørende dimensjonstabellene.

## Grunnleggende teknikker for fakatabeller

### Strukturen til en faktatabell

En faktatabell inneholder måltall produsert av en operasjonell hendelse måling i den virkelige verden. På det laveste kornet, tilsvarer en rad i faktatabellen en hendelse måling og vice versa. Dermed er den grunnleggende utformingen av en faktatabell helt og holdent basert på en fysisk aktivitet og er ikke påvirket av eventuelle rapporter som kan produseres. I tillegg til numeriske målinger, inneholder en faktatabell alltid en fremmednøkler for hver av de tilknyttede dimensjoner, samt valgfrie degenererte dimensjonsnøkler og dato-/tidsstempel. Faktatabeller er det primære målet for beregninger og dynamiske aggrigerte verdier som følge av spørringer.

### Additive, Semi-additive, og ikke-additive fakta

De numeriske målene i en faktatabell faller inn i tre kategorier. De mest fleksible og nyttige fakta er fullt additiv; de kan oppsummeres over en hvilken som helst av dimensjoner relatert til faktatabellen. De semi-additive måltallene kan summeres over noen av dimensjoner men ikke alle; saldobeløp er et typisk semi-additive fakta fordi de er summeres på tvers av alle dimensjoner unntatt tid. Til slutt er det noen måltall som er fullstendig ikke-additiv, slik som forholdstall. En god tilnærming for ikke-additive fakta er, om mulig, å lagre fullt additive komponenter av ikke-additivt måltall og summere disse komponentene i det endelige svaret satt før beregning av det endelig ikke-additive fakta. Denne siste beregningen er ofte gjort i BI laget eller OLAP kuben.

### NULL verdier i faktatabeller

Faktatabeller kan gjerne inneholde NULL-verder for måltall. Funksjoner som beregner aggrigerte verdier, som SUM, COUNT, MIN, MAX og AVG, håndterer alle NULL-verdier på den rette måten. NULL-verdier må imidlertid unngås i faktatabellens fremmednøklene til dimensjonstabeller, da dette vil føre til brudd med reglene for referanseintegriteten. Istedenfor NULL-verdier må de tilhørende dimensjonstabellene inneholde standard rader med surrigatnøkler som representerer ukjente verdier eller manglende verdier.

### Ulike fakta for samme måling

Hvis den samme målingen forekommer i flere faktatabeller, må man sørge for at de tekniske definisjonene denne faktaen er identiske hvis de skal sammenlignes eller beregnes sammen. Dersom de enkelte faktadefinisjonene er like bør faktaene ha identisk navnet, men hvis de er uforenlige så bør de ha ulike navn for å advare forretningsbrukere og BI-applikasjoner slik ad de er oppmerksom på forskjellen.

### Transaksjonsfaktatabeller

Hver rad i en transaksjonsfaktatabeller representerer en hendelsesmåling på et punkt i rom og tid. Atomiske transaksjonskorn faktatabeller er de mest dimensjonale og uttrykksfulle av faktatabeller fordi den robuste dimensjonaliteten muliggjør maksimal slicing og dicing av transaksjonsdata. Transaksjonsfaktatabeller kan være tette eller spredte fordi rader finnes bare hvis målingene finner sted. Disse faktatabeller inneholde alltid en fremmednøkkel for hver assosiert dimensjon og muligens også presise tidsangivelser og degenererte dimensjonsnøkler. De målte numeriske fakta må være konsistente med transaksjonen korn.

### Periodiske faktatabeller

Hver rad i en periodisk faktatabell oppsummerer flere målte hendelser over en standard periode som for eksempel en dag, uke eller måned. Detaljeringsnivået er perioden og ikke den enkelte transaksjonen. Periodisk faktatabeller inneholder ofte mange fakta fordi enhver målt hendelse som er i samsvar med faktatabellens detaljeringsnivå vil være aktuelle fakta. Disse faktatabellene er jevnt tette i forhold til femmednøkler fordi selv om det ikke har vært noen aktivitet i perioden så vil allikevel en rad bli lagt inn i faktatabellen med verdien 0 eller NULL for hver fakta.

### Akkumulerende øyeblikksbilde faktatabeller

Hver rad i en akkumulerende øyeblikksbilde faktatabell oppsummerer målte hendelser fra starten til slutten av en prosess med forhåndsbestemte steg. Arbeidsflytprosesser, for eksempel ordrebehandling eller prosessering av krav, som har et definert starttidspunkt, standardiserte steg i prosessen og et definert sluttpunkt kan modelleres med denne type faktatabell. Faktatabellen inneholder en dato fremmednøkkel i faktatabellen for hver kritisk milepæl i prosessen. De enkelte radene i en akkumulerende øyeblikksbilde faktatabell, eksempelvis en linje på en ordre, blir først opprettet når ordrelinje opprettes. Med fremskritt i arbeidsflyten, blir raden i den akkumulerende faktatabellen revidert og oppdatert. Denne kontinuerlige oppdateringen av rader i det akkumulerende øyeblikksbilde er unik blant de tre typene av faktatabeller. I tillegg til fremmednøkler for dato knyttet til hvert viktige steg i prosessen, vil en akkumulerende øyeblikksbilde faktatabell også inneholde fremmednøkler for andre dimensjoner og i tillegg til eventuelle degenererte dimensjoner. De inneholder ofte numeriske etterslepsmålingene som er konsistente med kornet, sammen med flagg for milepæler som oppnås.

### Faktaløse faktatabeller

Selv om de fleste målbare begivengheter fange resulterer i numeriske verdier, er det mulig at en begivenhet bare registrerer et sett med koblinger til dimensjonale entiteter som kommer sammen ved et gitt tidspunkt. For eksempel kan hendelsen av at en elev deltar på en forelesning på en gitt dag ikke ha et registrert numerisk faktum, men en faktarad med fremmednøkler til kalenderdag, student, lærer, lokasjon og klassen er veldefinerte. Likeledes, kommunikasjon med en kunde er en hendelse ofte uten noe numerisk målbar verdi. Faktaløse faktatabeller kan også brukes til å analysere hva som ikke skjedde. Disse spørringene har alltid to deler; en faktaløse dekningstabell som inneholder alle muligheter for hendelser som kan skje og en aktivitetstabell som inneholder de hendelsene som skjedde. Når aktivitetene trekes fra dekningen blir resultatet det sett av hendelser som ikke skjedde.

### Aggregerte faktatabeller eller kuber

Aggregerte faktatabeller er enkle numeriske samlinger av atomiske faktatabeller utelukkende for å forbedre ytelsen til spørringer. Disse aggregerte faktatabellene må være tilgjengelig for BI-laget samtidig som de atomiske faktatabellene slik at BI-verktøy fritt kan velge det riktige aggregerte nivå ved spørringer. Denne prosessen, kjent som aggregert navigering, må være åpent tilgjengelig slik at enhver som lager rapporter, spørreverktøy og BI-programmer kan høste de samme optimaliseringen i ytelse. Et riktig utformet sett av aggregater bør fungere seg som indekser i databasen for å akselerere ytelsen til spørringer, men ikke være direkte tilgjengelig for BI-applikasjoner og sluttbrukere. Aggregerte faktatabeller inneholde fremmednøkler til krympete ensartede dimensjoner i tillegg til aggregerte fakta beregnet ved å oppsummere målinger fra flere atomiske faktatabeller. Til slutt, er aggregerte OLAP-kuber med oppsummert målinger blir ofte bygget på samme måte som relasjonelle aggregeringer, men OLAP-kuber er ment å bli vist direkte av forretningsbrukerne.

### Konsolidert faktatabeller

Det er ofte fordelaktig å kombinere fakta fra flere prosesser sammen til en enkelt konsolidert faktatabell, hvis de kan uttrykkes med samme detaljeringsgrad. For eksempel, kan salgsfakta bli konsolidert med salgsprognoser i en enkelt faktatabell for å gjøre oppgaven med å analysere faktiske mot prognoser raskt og enkelt, i forhold til å måtte kjøre kompliserte krysskoblinger som benytter ulike faktatabeller. Konsoliderte faktatabeller legger en større byrde på ETL-behandling, men lette samtidig den analytiske byrden på BI-applikasjoner. De bør vurderes for beregninger på tvers av prosesser som ofte analyseres sammen.

## Grunnleggende teknikker for dimensjonstabeller

### Strukturen til en dimensjonstabell

Alle dimensjonstabeller har en enkelt primærnøkkelkolonne. Denne primærnøkkelen er implementert som en fremmednøkkel i eventuelle tilknyttede faktatabeller hvor en rad i dimensjonstabellen beskrivende konteksten for den aktuelle raden i faktatabellen.  
Dimensjonstabeller er vanligvis brede, flate denormaliserte tabeller med mange lav-kardinalitet tekstattributter. Mens operative koder og indikatorer kan behandles som attributter, inneholder de kraftigste dimensjonsattributter ordrike beskrivelser. Dimensjonstabellens attributter benyttes til å begrense og gruppere i spørringer og BI-applikasjoner. Beskrivende etiketter i rapporter er typisk atributtverdier fra dimensjonstabeller.

### Surrogatnøkler i dimensjonstabeller

En dimensjon tabellen er utformet med en kolonne som fungerer som en unik primærnøkkel. Denne primærnøkkelen kan ikke være den naturlige nøkkellen til det operative systemet fordi det vil være flere rader i dimensjonstabellen for denne naturlig nøkkelen når endringer spores over tid. I tillegg kan naturlige nøkler for en dimensjon lages av mer enn et kilde-systemet, og disse naturlige nøkler kan være inkompatible eller dårlig administrert. DW/BI-systemet må kreve kontroll over de viktigste nøklene til alle dimensjoner; istedenfor å bruke eksplisitte naturlige nøkler eller naturlige nøkler som inkorporerer en dato verdi, bør du opprette egne heltall primærnøkler for hver dimensjon. Disse surrogatene for dimensjonstabellene er enkle heltall, tildelt i stigende rekkefølge, og starter med verdien 1, hver gang en ny nøkkel er nødvendig. Dato dimensjonen er et unntatt fra surrogatnøkkelregelen; dette svært forutsigbare og stabiel dimensjonen kan bruke en mer meningsfylt primærnøkkel.

Hver dimensjonstabell bør i tillegg inneholde 2 ekstrarader som representerer "ingen verdi" og "ukjent verdi". Disse vil refereres til fra faktatabeller når fakta enten ikke har en relasjon til dimensjonen eller refererer til en dimensjonsentitet som ikke eksisterer i dimmensjonstabellen. Det siste forholdet må oppfates som en feilsituasjon som resultat av dårlig datakvalitet eller at ETL-prosessen oppdaterer innholdet i dimensjonstabellen sor sent (etter oppdatering av faktatabellen). Bruk de negative tallverdiene -1 og -2 som surrigatnøkler for disse tillegsradene i dimensjonatabellen.

### Naturlige, holdbare, og supernatural nøkler

Naturlige nøkler opprettet av operative kildesystemer er underlagt forretningsregler utenfor kontroll av DW/BI-systemet. For eksempel kan et ansatt nummer (naturlig nøkkel) endres dersom arbeidstakeren fratrer og deretter er ansettes igjen. Når datavarehuset ønsker å ha en enkelt nøkkel til denne arbeidstakeren må en ny nøkkel opprettes som er vedvarende og ikke endrer seg i denne situasjonen. Denne nøkkelen refereres ofte til som en holdbar naturlig nøkkel. De meste holdbare nøklene har et format som er uavhengig av den opprinnelige forretningsprosesser og dermed bør være et enkle heltall tildelt i sekvens begynner med verdien 1. Mens flere surrogatnøkler kan være forbundet med en ansatt over tid som deres profil endrer seg så vil den holdbar nøkkelen aldri endre seg.

### Drille ned

Drille ned er den mest grunnleggende måten data analyseres av forretningsbrukere. Drille ned betyr at man legger til en rad header til en eksisterende spørring; tittelen til den nye raden er en dimensjonsattributt lagt til GROUP BY-uttrykk i en SQL-spørring. Attributtet kan komme fra hvilken som helst dimensjon knyttet til faktatabellen i spørringen. Drille ned krever ikke definisjon av forhåndsbestemte hierarkier eller drill-down stier.

### Degenererte dimensjoner

Noen ganger er en dimensjon er definert uten innhold med bortsett fra sin primærnøkkel. For eksempel når en faktura har flere fakturalinjer så vil faktarader som representerer de enkelte fakturalinjene arve alle fremmednøklene til de beskrivende dimensjonene fra fakturaen, mens fakturaen er igjen uten noe unikt innhold. Men fakturanummeret er fortsatt en gyldig dimensjonsnøkkel for faktatabeller på linjenivå. Denne degenererte dimensjonen er plassert i faktatabellen med den eksplisitte erkjennelse at den ikke er forbundet med en dimensjonstabell. Degenererte dimensjoner er mest vanlig brukt sammen med transaksjoner og samler snapshot faktatabeller.

### Degenererte flate dimensjoner

Generelt må dimensjonale designere motstå fristelsen av å normalisering forårsaket av år med operative database design og i stedet denormalize de mange-til-en fast dybde hierarkier i separate attributter i en flat dimensjonsrad. Denormalisering av dimensjoner støtter dimensjonal modellering av tvillingmålene for enkelhet og hastighet.

### Flere hierarkier i dimensjoner

Mange dimensjoner inneholde mer enn ett naturlig hierarki. For eksempel kan dimensjoner for kalenderdato har en dag til uke til finans periode hierarki, så vel som en dag til måneder til år-hierarkiet. Beliggenhet intensive dimensjoner kan ha flere geografiske hierarkier. I alle disse tilfellene, kan de separate hierarkier grasiøst sameksistere i samme dimensjon tabellen.

### Flagg og indikatorer som dimensjonelle tekstattributter

Kryptiske forkortelser, ja/nei flagg og operasjonelle indikatorer bør i dimensjonstabeller suppleres med utfyllende ord som har betydning i seg selv uavhengig av kolonnetittel. Operasjonelle koder med innebygd mening i kodenverdien bør bli brutt ned med i hver del av koden til sin egen separate dimensjonattributter med en verdi som beskriver dens betydning.

### Null-attributter i dimensjoner

Dimensjonsattributter med null-verdier oppstår når en gitt rad i dimensjonstabellen ikke er fullstendig registrert, eller det er atributter som ikke er aktuelle for alle radene i dimensjonstabellen. I begge tilfeller anbefaler vi å benytte en beskrivende tekstverdi, som f.eks. Ukjent eller Ikke relevant istedenfor null-verdien. Null-verdier i dimensjonstabellen bør unngås fordi ulike databaser har ulik håndtering gruppering og begrensninger av null-verdier.

### Tidsdimensjonen

De fleste faktatabeller vil være knyttet til tidsdimensjon som gjør det mulig å navigering i faktatabellen med kjente datoer, måneder, regnskapsperioder og spesielle dager på kalenderen. Du vil ikke ønske å beregne påsken med en SQL-spørring, men heller slå den opp i en tidsdimensjon. Tidsdimensjonen vil typisk ha mange attributter som beskriver egenskaper; ukenummer, navn på måned, regnskapsperiode og indikator for offentlige helligdager. I stedet for en sekvensielt tildelt surrogat nøkkel vil primærnøkkelen til tidsdimensjonen være et heltall som representerer ÅÅÅÅMMDD, som er mer meningsfull og som legger til rette for partisjonering. Tidsdimensjonstabellen må i tillegg inneholde 2 spesielle rader som henholdsvis representere et ukjent tidspunkt og et tidspunkt som vil bli bestemt senere. Hvis en intelligent dato nøkkel benyttes bør filtrering og gruppering være basert på dimensjonsattributter og ikke på den intelligente nøkkelen.

Er det nødvendig med ytterligere presisjon av tidspunkt inkluderes et egen tidsstempel i faktatabellen. Dette tidsstempelet er en frittstående kolonneikke og ikke fremmednøkkel til en dimensjonstabell.

Hver rad i dimensjonstabellen **Dato Dimensjon** representerer en gitt dato. Fakta kan aggregeres i et hierarki bestående av dato, måned, kvartal og år.

|  |  |  |
| --- | --- | --- |
| **Kolonne** | **Datatype** | **Beskrivelse** |
| Dato Key | int\* | Nøkkelverdi som identifiserer en gitt dag i datavarehuset, formatert som ÅÅÅÅMMDD f.eks. 20071231 |
| Måned Key | int\* | Fremmednøkkel til Måned Dimensjon som angir hvilken måned som dagen tilhører, formatert som ÅÅÅÅMM f.eks. 200712 |
| Dato | date\* | Dag som dato med format ÅÅÅÅ-MM-DD. Eksempelverdi: 2007-12-31 |
| År | smallint\* | Årstall som heltall [ÅÅÅÅ]. Eksempelverdi: 2007 |
| Kvartal nummer | tinyint\* | Kvartal som heltall [1-4]. Eksempelverdi: 4 |
| Kvartal med år nummer | smallint\* | Årstall og kvartal som heltall ÅÅÅÅ[1-4]. Eksempelverdi: 20074 |
| Måned nummer | tinyint\* | Måned som heltall [1-12]. Eksempelverdi: 12 |
| Måned med år nummer | smallint\* | Årstall og måned som heltall ÅÅÅÅ[01-12]. Eksempelverdi: 200712 |
| Dag i måned nummer | tinyint\* | Dag i måneden som heltall [1-31]. Eksempelverdi: 31 |
| Uke | tinyint\* | Ukenummer som heltall [1-53]. Eksempelverdi: 52 |
| Uke med år nummer | smallint\* | Årstall og ukenummer som heltall ÅÅÅÅ[01-53]. Eksempelverdi: 200752 |
| Ukedag nummer | tinyint\* | Ukedag som heltall [1-7]. Man=1, Tir=2, Ons=3, Tor=4, Fre=5, Lør=6, Søn=7 |
| Kvartal | char(2)\* | Kvartal formatert som K[1-4]. Eksempelverdi: K4 |
| Kvartal med år | char(6)\* | Årstall og kvartal, formatert som ÅÅÅÅK[1-4]. Eksempelverdi: 2007K4 |
| Måned | varchar(9)\* | Fullt navn på måned. Eksempelverdi: desember |
| Mnd | char(3)\* | Forkortet navn (3 bokstaver) på måned. Eksempelverdi: des |
| Måned med år | char(6)\* | Måned med årstall, formatert som [jan-des].ÅÅ. Eksempelverdi: des.07 |
| Ukedag | varchar(7)\* | Fullt navn på ukedag. Eksempelverdi: tirsdag |
| Dag | char(3)\* | Forkortet navn (3 bokstaver) på ukedag. Eksempelverdi: tir |

*Dimensjonstabellen Dato Dimensjon (skript for opprettelse av tabellen finner du her)*

Hvis sluttbruker ønsker å filtrere eller gruppere i forhold til tidspunkt på døgnet, som f.eks. gruppering av timer eller skiftnummer, vil det imidlertid være lurt å implementere en egen tid-på-døgnet dimensjon og en fremmednøkkel til denne i faktatabellen.

Dersom en virksomhet opererer med noen fakta som er månedsbaserte og andre fakta som en datobaserte kan det være hensiktsmessig å implementere tidsdimensjonen som 2 [samsvarende dimensjoner](http://www.bedreinnsikt.no/innhold/samsvarende-dimensjoner); en månedsdimensjon og en datodimensjon.

Hver rad i dimensjonstabellen **Måned Dimensjon** representerer en gitt måned. Fakta kan aggregeres i et hierarki bestående av måned, kvartal og år. En månedlig periode vil her bli satt til datoen for den første dagen i måneden.

|  |  |  |
| --- | --- | --- |
| **Kolonne** | **Datatype** | **Beskrivelse** |
| MånedKey | int\* | Nøkkelverdi som identifiserer en gitt måned i datavarehuset, formatert som ÅÅÅÅMM. Eksempelverdi: 200712 |
| Første dag Dato Key | int\* | Fremmednøkkel til Dag Dimensjon som angir hvilken dag som representerer måneden, formatert som ÅÅÅÅMMDD. Eksempelverdi: 20071231 |
| År | smallint\* | Årstall som heltall [ÅÅÅÅ]. Eksempelverdi: 2007 |
| Kvartal nummer | tinyint\* | Kvartal som heltall [1-4]. Eksempelverdi: 4 |
| Kvartal med år nummer | smallint\* | Årstall og kvartal som heltall ÅÅÅÅ[1-4]. Eksempelverdi: 20074 |
| Måned nummer | tinyint\* | Måned som heltall [1-12]. Eksempelverdi: 12 |
| Måned med år nummer | smallint\* | Årstall og måned som heltall ÅÅÅÅ[01-12]. Eksempelverdi: 200712 |
| Kvartal | char(2)\* | Kvartal formatert som K[1-4]. Eksempelverdi: K4 |
| Kvartal med år | char(6)\* | Årstall og kvartal, formatert som ÅÅÅÅK[1-4]. Eksempelverdi: 2007K4 |
| Måned | varchar(9)\* | Fullt navn på måned. Eksempelverdi: desember |
| Mnd | char(3)\* | Forkortet navn (3 bokstaver) på måned. Eksempelverdi: des |
| Måned med år | char(6)\* | Måned med årstall, formatert som [jan-des].ÅÅ. Eksempelverdi: des.07 |

*Dimensjonstabellen Måned Dimensjon (skript for opprettelse av tabellen finner du her)*

### Dimensjon for roller

Den samme fysiske dimensjonen kan refereres til flere ganger av samme faktatabell, hvor hver referansekobling representerer en logisk rolle mot dimensjonen. For eksempel kan en faktatabell har flere datoer, hvor hver av disse blir representert av et fremmednøkkel til dato-dimensjonen. Det er viktig at hver sekundærnøkkel refererer til et eget perspektiv på det tidspunkt dimensjon slik at referansene er uavhengige. Disse ulike dimensjonsperspektivene (med unike atributt-kolonnenavn) kalles roller.

### Oppsamlingsdimensjoner

Transaksjonelle forretningsprosesser produserer vanligvis en rekke diverse flagg og indikatorer med lav kardinalitet. Istedenfor å lage egne dimensjonstabeller for hvert flagg og hver attributt kan du lage en felles dimensjon for alle disse. Denne dimensjon, ofte betegnet som en transaksjon profildimensjon i et skjema, behøver ikke å være det kartesiske produkt av alle mulige verdier for disse atributtene, men trenger kun å inneholde kombinasjon av de verdiene som faktisk opptrer i kildedataene.

### Snøfnuggdimensjon

Når en hierarkisk relasjon i en dimensjonstabell er normalisert, opptrer attributter med lav kardinalitet som sekundære tabeller som er koblet til basisdimensjonstabellen gjennom en attributtnøkkel. Når denne prosessen gjentas for alle hierarkiene i dimensjonenstabellen så er en karakteristisk struktur med flere nivå opprettet som kalles et snøfnugg. Selv om snøfnugg representerer hierarkiske data korrekt bør du allikevel unngå snøfnugg fordi det er vanskelig for forretningsbrukere å forstå og navigere snøfnugg. De kan også ha negativ innvirkning på ytelsen til spørringer. En flat denormalized dimensjonstabellen inneholder nøyaktig samme informasjon som en snøfnuggdimensjon.

### Støttedimensjoner

En dimensjon kan inneholde en referanse til en annen dimensjonstabell. For eksempel kan en bankkonto-dimensjon referere til en egen dimensjon som representerer datoen som kontoen ble opprettet. Slike sekundære dimensjonsreferanser kalles støttedimensjoner. Støttedimensjoner er tillatt, men bør brukes med måte. I de fleste tilfeller bør sammenhengene mellom dimensjonene bli flyttet til en faktatabell, der begge dimensjoner er representert som egne fremmednøkler.

## Integrasjon via overenstemte dimensjoner

### Samsvarende dimensjoner

Dimensjoner er samsvarende når attributter i dimensjonstabellene har samme kolonnenavn med likeformet innholdet. Informasjon fra ulike faktatabeller kan kombineres i en og samme rapport ved bruk av samsvarende dimensjonsattributter knyttet til faktatabellene. Når en samsvarende attributtet benyttes som radtittel (det vil si som grupperingskolonnen i SQL-spørringen) kan resultatene fra de enkelte faktatabellene stilles opp på de samme radene i en "drill-over" rapport. Dette er kjernen i integrasjonen til en BI-løsningen. Samsvarende dimensjoner defineres i tett samarbeid med representanter for virksomhetens dataforvaltning. Disse dimensjonene sikrer analytisk konsistens fordi de gjenbrukes på tvers av faktatabeller, samtidig som de reduserer fremtidige utviklingskostnader da hjulet ikke trengs å finnes opp på nytt.

Et godt eksempel er [tidsdimensjonen](http://www.bedreinnsikt.no/innhold/tidsdimensjonen) til et datavarehus når den implementeres som to samsvarende dimensjoner; en for måned og en for dato. På denne måten kan månedsbaserte fakta som f.eks. budsjetttall stilles opp i samme rapport som datobaserte fakta som f.eks. beløp postert i regnskapet.

### Reduserte opprullingsdimensjoner

Reduserte dimensjoner er opprullingsdimensjoner som er et delmengde av rader og/eller kolonner til en baisdimensjon. Reduserte opprullingsdimensjoner er nødvendig for å konstruere aggregerte faktatabeller. De er også nødvendig for forretningsprosesser som naturlig fanger data på et høyere detaljnivå, for eksempel en prognose for måned og merkevare (i stedet for den mer atomiske dato og produkter knyttet til salgsdata). Et annet tilfelle av overenstemte dimensjoner som understilles oppstår når to dimensjoner er på samme detaljnivå, men man representerer bare en delmengde av radene.

### Drille på tvers

Drilling på tvers betyr ganske enkelt å lage egne spørringer mot to eller flere faktatabeller hvor radtittelen for hver spørring består av overensstemmende attributter. Resultatsettet fra de to spørringene justeres ved å utføre en sortert flettingsoperasjon på de atributtene i dimensjonene med like radtitler. Leverandører av BI-verktøy referere til denne funksjonaliteten ved ulike navn, blant annet stitch og multipass query.

### Verdikjede

En verdikjede identifiserer den naturlige flyten av organisasjonens primære forretningsprosesser. For eksempel kan verdikjeden til en forhandler består av innkjøp til lagerstyring til salg. Verdikjeden til en hovedbok kan bestå av budsjettering til forpliktelser til utbetalinger. Operasjonelle kildesystemer produserer vanligvis transaksjoner eller øyeblikksbilder på hvert trinn i verdikjeden. Fordi hver prosess produserer unike beregninger i unike tidsintervaller med unike detaljnivå og dimensjonalitet så vil hver prosess vanligvis omspenne minst en atomisk faktatabell.

### Bussarkitektur for datavarehus

Bussarkitekturen til bedriftens datavarehus muliggjør en trinnvis tilnærming til å bygge virksomheten DW/BI-systemet. Denne arkitekturen dekomponerer DW/BI planprosessen i håndterbare biter ved å fokusere på forretningsprosesser samtidig som den sikrer integrasjon via standardiserte overensstemmende dimensjoner som gjenbrukes på tvers av prosesser. Dette gir et arkitektonisk rammeverk samtidig som det oppmuntre til håndterbar og smidig implementeringer av bedriftens datavarehus. Bussarkitekturen er teknologi og database plattformuavhengig; både relasjonsbaserte og OLAP-baserte dimensjonale strukturer kan utnytte arkitekturen.

### Bussmatriser for datavarehus

Bussmatrisen til bedriften datavarehus er et essensielle verktøyet for å designe og kommunisere bussarkitektur til bedriftens datavarehus . Radene i matrisen er forretningsprosesser og kolonnene er dimensjoner. De skyggelagte cellene i matrisen indikere at en dimensjon er assosiert med en gitt forretningsprosess. Designteamet skanner hver rad for å teste om en kandidatdimensjon er godt definert for forretningsprosesser og skanner hver kolonne for å se hvor en dimensjon bør være overensstemmende på tvers av flere forretningsprosesser. Foruten de tekniske designhensyn, blir bussmatrisen brukt bedriftsledere som input for å prioritere DW/BI prosjekter hvor utviklingsteamet vil implementere en rad i matrisen om gangen.

Den detaljerte bussmatrisen for implementering er en mer detaljert bussmatrise hvor hver forretningsprosessrad er blitt utvidet til å vise de spesifikke faktatabellene eller OLAP-kubene. På dette detaljnivået det endelige detaljnivået og en liste ove alle fakta dokumenteres.

### Mulighet/interessent matrise

Etter at radene i bussmatriser for bedriftens datavarehus er blitt identifisert kan du lage et utkast til en annen matrise ved å erstatte dimensjonskolonner med forretningsfunksjoner, for eksempel markedsføring, salg og økonomi, og deretter skyggelegging matrisecellene for å indikere hvilke forretningsfunksjoner er interessert i hvilke forretningsprosesser. Mulighet/ interessent matrisen bidrar til å identifisere hvilke forretningsgrupper bør inviteres til samarbeidøkter rundt design for hver prosess-sentrisk rad.

## Teknikker for sakteendrede dimensioner

### Type 0: Beholde originalen

Med sakte endrede dimensjonen av typen 0 vil attributtverdien aldri oppdateres, slik at fakta alltid er gruppert etter den opprinnelige verdien. Det er hensiktsmessig å merke atributter av type 0 med "original" som f.eks. en kundes opprinnelige kreditt scorer eller en varig identifikator som personnummer. Det gjelder også for de fleste attributtene i en dato dimensjon.

### Type 1: Overskriv

Med sakte endrede dimensjonen av typen 1 vil attributtverdien overskrives med nye verdier slik at attributter alltid gjenspeiler den siste tilordningen. Dermed vil denne teknikken ødelegge all historikk. Selv om denne tilnærmingen er enkel å implementere og ikke skaper ekstra rader i dimensjonstabellen må du være påpasselig med at aggregerte faktatabeller og OLAP kuber som er berørt av denne endringen blir beregnes på nytt.

### Type 2: Legg til ny rad

Med sakte endrede dimensjonen av typen 2 vil nye rader legges til dimensjontabellen når med de oppdaterte attributtverdier. Dette krever generalisering av primærnøkkelen for dimensjonen utover den naturlig eller holdbare nøkkelen fordi det potensielt vil være flere rader som beskriver hvert medlem. Når en ny rad er opprettet for et dimensjonsmedlem, blir en ny primær surrogatnøkkelen tildelt og brukt som en fremmednøkkel i alle faktatabeller fra det øyeblikket av oppdateringen inntil en senere endring skaper en ny dimensjonsnøkkel og oppdatert dimensjonsrad.

Et minimum av ytterligere tre kolonner bør legges til dimensjonen rad med type 2 endringer:

1. rad for effektiveringsdato eller dato/klokkeslett-stempel
2. rad for utløpsdato eller dato/klokkeslett-stempel
3. indikator for gjeldende rad

## Teknikker for dimensjonshierarki

### Posisjonelle hierarki med fast dybde

Et hierarki med fast dybde er en serie med mange-til-en-forhold, som for eksempel produkt til merkevare til kategori til avdelingen. Når et hierarki med fast dybde er definert og man er blitt enige om navnene til de enkelte hierarkinivåene, bør de enkelte hierarkinivåene vises som egne posisjonsattributter i dimensjonstabellen. Et hierarki med fast dybde er den klart enkleste å forstå og navigere seg i så lenge de ovennevnte kriteriene er oppfylt. Den tilbyr også forutsigbar og rask ytelse på spørringer. Når hierarkiet ikke er en serie av mange-til-én-forhold eller antallet nivåer kan varierer slik at man ikke blir enige om navn på nivåene så må en teknitt for hierarki med fast dybde benyttes.

### Hierarkier med litt variabel dybde

Hierarkier med litt variabel dybde har ikke et fast antall nivåer, men variasjonen i dybde er liten. Geografiske hierarkier vil ofte varierer i dybde med fra kanskje tre nivåer til seks nivåer. Istedenfor enn å bruke den komplekse teknikken for uforutsigbart variable hierarkier, kan du tvinge et litt fillete hierarkier til en fast dybde posisjonell design med egen dimensjon attributter for det maksimale antall nivåer, og deretter fylle attributtverdien basert på regler fra virksomheten.

### Hierarkier med variabel dybde

Taggede hierarkier med ubestemmelig dybde er vanskelig å modellere og utforme spørringer for i en relasjonsdatabase. Selv om SQL-utvidelser og OLAP spørrespråk gir en viss støtte for rekursive foreldre/barn-relasjoner så har disse tilnærmingene begrensninger. Med SQL-utvidelser kan alternative taggede hierarkier ikke omformes når spørringen utføres, delte eierstrukturer støttet ikke og tidsvarierende taggede hierarkier støttes ikke. Alle disse begrensningene kan løses i relasjonsdatabaser ved å modellere et fillete hierarki med en spesialkonstruert brotabellen. Denne brotabellen inneholder en rad for hver mulig sti i det fillete hierarkiet og gjør at alle former for traverseringer av hierarkiet kan oppnås med standard SQL istedenfor å bruke spesielle språkutvidelser.

Bruken av en brotabell for taggede hierarkier med variabel dybde kan unngås ved å implementere en stistrengattributt i dimensjonen. For hver rad i dimensjonstabellen inneholder stistrengen en spesielt kodet tekststreng som inneholder den fullstendige beskrivelse av stien fra den øverste node i hierarkiet ned til noden beskrevet av den aktuelle dimensjonsraden. Mange av de vanlige analyse hierarkiforespørslene kan dermed bli behandlet med standard SQL uten å ty til språkutvidelser. Stistreng tilnærmingen vil imidlertid ikke muliggjøre rask substitusjon av alternative hierarkier eller hierarkier med delt eierskap. Stistreng tilnærming kan også være sårbare for strukturelle endringer i taggede hierarki som kan tvinge hele hierarkiet til å bli bygget opp på nytt.

# Arkitektur for ETL-prosessen

Ekstrahere, transformere og laste (ETL) systemet bruker en uforholdsmessig stor andel av tiden og innsatsen som kreves for å bygge et datavarehus og business intelligence (DW/BI ) miljø . Utvikling av ETL-systemet er utfordrende fordi så mange utenforliggende begrensninger legge press på sin designet; forretningsmessige krav, kildedatarealiteter, budsjett, prosessering vindu, og ferdigheter til tilgjengelig personell. Likevel kan det være vanskelig å forstå hvorfor ETL-systemet er så komplisert og ressurskrevende.

Alle forstår de tre bokstavene; Du få data ut av sin opprinnelige kilde plassering (E) , du gjør noe med det (T) og deretter laster du det (L) til et endelig sett med tabeller for forretningsbrukere til å spørre mot. På spørsmål om den beste måten å designe og bygge ETL-systemet sier mange designere; "Vel , det kommer an på." Det kommer an på kilden, det kommer an på begrensninger til dataene, det kommer an på skriptspråkene og ETL-verktøyene som er tilgjengelig, det kommer an på personalets kompetanse og det kommer an på BI-verktøyene. Men "det kommer an på" svar er farlig fordi det blir en unnskyldning for å ta en ustrukturert tilnærming til å utvikle en ETL-system, som i verre fall resulterer i et udifferensiert spaghettirot av tabeller, moduler, prosesser, prosedyrer, triggere, varslinger og jobbplaner. Denne "kreativ" designtilnærming bør ikke tolereres.

Med visdom og etterpåklokskap fra tusenvis av dimensjonale datavarehus har et sett med ETL beste praksis dukket opp. Nøye vurdering av disse beste praksisene har avdekket 34 delsystemer som kreves i bakrommet til nesten alle dimensjonale datavarehus. Kimball konsernet har organisert disse 34 delsystemer av ETL-arkitekturen inn i kategorier:

* Tre delsystemer fokuserer på å ekstrahere data ut fra kildesystemene.
* Fem delsystemer håndtere verdiøkende vasking og konsolidering, inkludert dimensjonale strukturer for å overvåke kvalitetsfeil.
* Tretten delsystemer levere data som dimensjonale strukturer til den endelige BI-laget, slik som et delsystem for å implementere teknikker for langsomt skiftende dimensjoner.
* Tretten delsystemer for å styre ETL produksjonsmiljøet.

# Oversettelse av engelske begreper

Prøver du å finne norsk oversettelse for ord og uttrykk relater til datavarehus og business intelligence? Da har du kommet til rett sted! Under finner du en omfattene liste av begreper som som vi har oversatt til norsk.

* **Accelerate Query Performance** - øke ytelsen til spørringer
* **Accumulating Snapshot Fact Table \*** - akkumulerende øyeblikksfaktatabell, akkumulerende øyeblikksbildefaktatabell
* **Actuals versus Forecasts \*** - faktiske mot prognoserte, faktiske mot prognoser, sammenligning av faktiske verdier opp mot prognoseverdier
* **Additive Fact \*** - additive fakta, additative fakta
* **Add New Row** - legge til nye rad
* **Aggregate Fact Table** - aggregert faktatabell, faktatabell med oppsummerte fakta
* **Aggregate Function \*** - aggregatfunksjon, aggregert funksjon
* **Aggregate Level** - aggregeringsnivå, aggregert nivå
* **Aggregate Navigation \*** - aggregert navigasjon
* **Aggregate OLAP Cube \*** - aggregerte OLAP-kuber
* **Aggregated Data** - aggregerte data
* **Aggregation** - aggregering
* **Agile Implementation** - smidig implementering
* **Analytic Consistency \*** - konsistente analyseresultater
* **Application Processing** - behandling av søknader
* **Architectural Framework** - arkitektonisk rammeverk, rammeverk for arkitekturen
* **Associated Dimension \*** - assosiert dimensjon, tilknyttet dimensjon
* **Associated Dimension Table \*** - assosiert dimensjonstabell, tilknyttet dimensjonstabell
* **Atomic Data \*** - atomiske data
* **Atomic Fact Table \*** - atomisk faktatabell
* **Atomic Transaction Grain Fact Table \*** - transaksjonsfaktatabell med atomisk detaljgrad
* **Base Dimension Table \*** - grunnleggende dimensjonstabell
* **BI Application \*** - BI-applikasjon
* **BI Layer \*** - BI-laget
* **Bridge Table \*** - brotabell, brobyggingstabell
* **Business Analysis** - forretningsanalyse
* **Business Function \*** - forretningsfunksjon
* **Business Group** - faggruppe
* **Business Process** - forretningsprosess
* **Business Process Event** - hendelse i forretningsprosessen
* **Business Rule** - forretningsregel
* **Business User \*** - sluttbruker
* **Calendar Date Dimension \*** - tidsdimmensjon, dimensjon for kalenderdato, datodimensjon
* **Calendar Date Dimension \*** - dimensjon for kalenderdato
* **Capture Data \*** - fange opp data, samle inn data, registrere data
* **Cartesian Product of** - kartesisk produkt av
* **Claim Processing** - behandling av krav
* **Collaborative Design Session \*** - designøkter i samarbeid
* **Collaborative Workshop \*** - sammarbeidende arbeidsmøte
* **Conformed Dimension Attributes** - samsvarende dimensjonsatributter
* **Conformed Dimensions** - samsvarende dimensjoner
* **Conformed Facts** - samsvarende fakta
* **Consolidated Fact Table** - konsolidert faktatabell
* **Constraints and Grouping \*** - begrensninger og gruppering
* **Cross-Process Metrics \*** - cross-prosess beregninger?
* **Cryptic Abbreviation** - kryptiske forkortelse
* **Cube** - datakube
* **Current Row Indicator \*** - indikator gjeldende rad, gjeldende rad indikator
* **Database Management System \*** - databasehåndteringssystem
* **Database Platform** - databaseplattform
* **Data Governance** - dataforvaltning
* **Data Mart \*** - temavarehus, datamarked
* **Data Profiling** - dataprofilering
* **Data Quality** - datakvalitet
* **Data Source** - datakilde
* **Date Stamp** - datostempel
* **Data Steward** - dataansvarlig
* **Data Warehouse** - datavarehus
* **Data Warehouse Methodology** - datavarehusmetodikk
* **Date/time Stamp \*** - dato/tidsstempel
* **Day Part Grouping \*** - døgngruppering
* **Declared Grain \*** - xxxxx
* **Degenerated Dimension Key \*** - degenerert dimensjonsnøkkel
* **Degenerate Dimensions \*** - degenererte dimensjoner
* **Denormalized Table** - denormalisert tabell
* **Denormalized Flattened Dimension \*** - denormalisert flat dimensjon
* **Dense Fact Table \*** - tett faktatabell
* **Descriptive Context** - beskrivende kontekst
* **Descriptive Labels \*** - beskrivende etiketter (titler?)
* **Descriptive String** - beskrivende tekststreng
* **Dimension** - dimensjon
* **Dimension Attribute \*** - dimensjonsattributt
* **Dimensional Design Process \*** - dimensjonell designprosess
* **Dimension Hierarchy** - dimensjonshierarki
* **Dimensional Modeling** - dimensjonsmodellering
* **Dimension Surrogate Key \*** - surrogatnøkkelen til dimensjonen, dimensjonens konstruerte nøkkel
* **Dimension Table** - dimensjonstabell
* **Detailed Implementation Bus Matrix \*** - detaljert bussmatrise for implementasjon
* **Domain Contents \*** - domeneinnhold
* **Domain Value \*** - domeneverdi, forretningsområdeverdi
* **Drill-across \*** - drilling på tvers
* **Drill-Across Report \*** - drill-over rapport
* **Drill-down \*** - datadrilling
* **Drill-Down Path \*** - drill-down sti
* **Drilling Down \*** - drille ned, detaljoversikt, forfine en av dimensjonene
* **Drill-up \*** - oppsummering
* **Durable Key \*** - slitesterk nøkkel
* **Effective Date \*** - effektiveringsdato, gyldig fra dato
* **End Point \*** - endepunkt, slutten
* **Enterprise Data Warehouse Bus Architecture \*** - bussarkitektur til bedriftens datavarehus
* **Enterprise Data Warehouse Bus Matrix \*** - bussmatrise til bedriftens datavarehus
* **Enterprise Information Management \*** - informasjonsforvaltning
* **ETL (Extract-Transform-Load)** - ekstrahere, transformere og laste
* **ETL Process** - ETL-prosess
* **Expiration Date \*** - utløpsdato, gyldig til dato
* **Event** - hendelse
* **Fact** - fakta
* **Fact Entity** - faktaentitet
* **Factless Fact Table** - faktaløse faktatabell
* **Fact Table** - faktatabell
* **Filtering the Facts** - filtrering av fakta
* **Fixed Depth Hierarchy** - hierarki med fast dybde
* **Fixed Depth Positional Hierarchy \*** - posisjonellt hierarki med fast dybde
* **Fixed Number of Levels \*** - fast antall nivå
* **Flattened Denormalized Dimension Table \*** - denormalisert flat dimensjonstabell
* **Flattened Dimension Row \*** - utflatet dimensjonsrad
* **Forecast by Month \*** - månedlig prognose, prognose etter måned
* **Foreign Key** - fremmednøkkel
* **Fscal Period \*v** - fscal periode
* **Fullment Processing \*** - processing av innfrielse
* **Geographic Hierarchy** - geografisk hierarki
* **Grain \*** - detaljeringsnivå, detaljgrad, detaljeringsgrad
* **Grain Statement \*** - bestemmelse av detaljgrad
* **Grain Declaration \*** - deklarasjon av detaljgrad, deklarasjon av detaljeringsgrad
* **Grouping the Facts** - gruppering av fakta
* **Hierarchical Relationship** - hierarkisk relasjon
* **Hierarchy Level** - hierarkinivå, nivå i hierarkiet, hierarkisk nivå
* **Higher Level of Granularity \*** - høyere nivå av granularitet
* **Indeterminate Depth \*** - ubestemmelige dybde, uforutsigbar dybde
* **Intermediate Steps** - mellomsteg
* **Junk Dimension \*** - skrotdimmensjon
* **Lag Measurement \*** - måling av forsinkelse, etterslepsmåling
* **Level of Detail** - detaljnivå
* **Low-Cardinality** - lav kardinalitet
* **Many-to-Many** - mange-til-mange
* **Many-to-One** - mange-til-en
* **Many-to-One Relationship \*** - mange-til-en relasjon, mange-til-en forhold
* **Measurement** - måling
* **Measurement Event \*** - målbar hendelser
* **Metadata** - metadata
* **Milestone** - milepæl
* **Milestone Completion Counter \*** - teller for ferdigstilt milepæl
* **Multiple Hierarchies in Dimension \*** - flere hierarkier i dimensjonen
* **Natural Hierarchy \*** - naturlig hierarki
* **Natural Key \*** - naturlige nøkkel
* **Non-Additive Fact \*** - ikke-additive fakta
* **Normalized** - normalisert
* **Not Applicable** - ikke relevant
* **Null Attribute in Dimension \*** - Null-attributter i dimensjon, dimensjonell null-atributt
* **Null Foreign Key \*** - fremmednøkkel med NULL-verdier
* **Null-Valued Measurement \*** - måling med NULL-verdier
* **Numeric Measure \*** - numerisk måling
* **OLTP (On-Line Transactional Processing** - driftssystem, transaksjonssystem
* **OLAP (On-Line Analytical Processing)** - beslutningsstøttesystem
* **OLAP Tool** - OLAP-verktøy
* **Operational Measurement Event \*** - målbar operasjonell begivenhet
* **Opportunity/Stakeholder Matrix \*** - mulighet-/interessentmatrise
* **Optional** - valgfritt
* **Operational Code \*** - operasjonell kode, driftskode, håndteringskode, funksjonskontroll kode
* **Operational Database** - produksjonsdatabase, virksomhetens operasjonelle database
* **Operational Indicator \*** - operasjonell indikator, virksomhetsindikator
* **Outrigger Dimensions \*** - utriggede dimensjoner
* **Overwrite \*** - overskriv, erstatt
* **Query Performance \*** - spørreytelse
* **Query Time \*** - spørretidspunkt, tidspunkt for kjøring av spørring
* **Query Tool** - spørreverktøy
* **Pathstring Attribute \*** - stistreng attributt
* **Periodic Snapshot Fact Table \*** - periodisk øyeblikksbildefaktatabell
* **Performance Metric \*** - ytelsesmåling
* **Performance Tuning \*** - forbedring av ytelse
* **Pipeline Process \*** - faseinndelt prosess
* **Predictable Steps** - forutsigbare steg
* **Precise Time Stamp \*** - presis tidsangivelse
* **Predetermined Hierarchy** - forhåndsbestemt hierarki
* **Predictable and Fast Query Performance \*** - forutsigbar og rask spørreytelse
* **Primary Surrogate Key** - primær surrogatnøkkel
* **Primary Key** - primærnøkkel
* **Populate \*** - fylle ut, legge inn, registrere
* **Positional Attribute \*** - posisjonelle attributt
* **Process Measurement Event \*** - målbar prosesshendelser
* **Process Step \*** - steg i prosessen
* **Ragged Hierarchy \*** - fillete hierarki, tagge-de hierarki
* **Range in Depth \*** - dybdeforskjell, variasjon i dybde
* **Recursive Parent/Child Relationship \*** - rekursiv foreldre-barn relasjon
* **Referential Integrity Violation \*** - brudd i referanseintegritet
* **Reporting Cube** - rapporteringskube
* **Report Writer \*** - rapporten forfatter
* **Retain Original** - behold originalen
* **Reuse Across Processes \*** - gjenbruk på tvers av prosesser
* **Role-Playing Dimension \*** - rolledimensjon
* **Rolled-up \*** - rollup, aggregere data
* **Rolled-up Summary \*** - rullet opp sammendrag
* **Sales Forecast** - salgsprognose
* **Secondary Table** - sekundærtabell
* **Semi-Additive Fact \*** - semi-additive fakta
* **Shared Ownership Structure \*** - delt eierskapsstrukturer, struktur for delt eierskap, felles eierskapsstrukturer
* **Shift Number \*** - skiftnummer
* **Shrunken Rollup Dimensions \*** - forminsker/krympet opprullingsdimensjon
* **Slice and dice \*** - akkumulere langs noen dimensjoner og eksponere andre
* **Slightly Ragged Hierarchy \*** - litt fillete hierarki, noe fillete hierarki
* **Slightly Variable Depth Hierarchy \*** - hierarki med noe variabel dybde
* **Slowly Changing Dimension \*** - sakteendrede dimensjon, tregt endrede dimensjon
* **Snapshot Fact Table \*** - øyeblikksbilde faktatabell, øyeblikksfaktatabell
* **Snowflake \*** - snøflak, snøflakskjema
* **Snowflake Dimension \*** - snøflakdimensjon
* **Source System** - kildesystem
* **Sparse Fact Table** - glissen faktatabell
* **SQL Extension** - SQL-utvidelse
* **SQL Query** - SQL-spørringer
* **Star Schema** - stjerneskjema
* **Start Point \*** - startpunkt, begynnelse
* **Subset of Columns** - delsett av kolonner
* **Subset of Rows** - delsett av rader
* **Summed Across Dimensions** - summert på tvers av dimensjoner
* **Summarize** - oppsummere
* **Summarized Measures** - oppsummert målinger
* **Textual Dimension Attributes \*** - Tekstlige dimensjonsattributter
* **Time-of-Day \*** - tid på døgnet, tid på dagen, tidspunkt, dele opp en dag i hele timer
* **Time Stamp** - tidsstempel
* **Transaction Data** - transaksjonsdata
* **Transaction Fact Table** - transaksjonsfaktatabell
* **Transaction Profile Dimension \*** - transaksjonsprofildimensjon
* **Transaction Snapshot Fact Table \*** - transaksjonens øyeblikksbilde faktatabell, øyeblikksfaktatabell for transaksjoner
* **True/False Flags \*** - ja/nei flagg
* **Uniformly Dense Fact Table \*** - jevnt tett faktatabell
* **User Queries** - brukerspørsmål
* **Value Chain** - verdikjede
* **Variable Depth Hierarchy** - hierarki med variabel dybde
* **Workfow Process** - arbeidsflytprosess

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

* **Data Mining** - datagruvedrift
* **?** - datakilde
* **?** - strategisk beslutningv
* **?** - design og implementasjon av datavarehus
* **?** - SQL-analysefunksjon
* **?** - en samling teknikker for å «oppdage mønstre» i store datasett
* **?** - beslutningsstøttesystemer
* **?** - ledelsen i bedriften
* **?** - databasesystem
* **?** - virksomhetskritisk
* **?** - daglig drift av en virksomhet
* **?** - kontinuerlig oppdatert
* **?** - historiske data
* **?** - to sider av en virksomhet − den operasjonelle og den strategiske
* **?** - informasjonssystem
* **?** - data er aggregert over tid
* **?** - kontorpersonell
* **?** - saksbehandlere
* **?** - selgere
* **?** - operatører
* **Throughput** - høy gjennomstrømning
* **?** - ledere
* **?** - ad hoc-spørringer
* **?** - data fra eksterne datakilder
* **?** - periodisk aggregert
* **?** - verktøy for analyse og rapportering
* **?** - strategiske valg
* **?** - design og implementasjon
* **?** - datamodell
* **?** - sammensatt primærnøkkel
* **?** - kvantitative (numeriske) data
* **?** - gruppering og bruk av mengdefunksjoner
* **?** - dimensjonstabell inneholder attributter
* **?** - dimensjonsentitet
* **?** - effektivitetsgevinst
* **?** - datavarehuset bør bygges opp
* **?** - utviklingsstrategi
* **?** - opplasting av data
* **?** - vasking av data
* **?** - økt datakvalitet
* **?** - datakonvertering
* **?** - indekseringsteknikker
* **?** - bitmap-indekser
* **?** - indekser
* **?** - analyse og rapportering
* **?** - grafisk brukergrensesnitt
* **Slicing** - slicing
* **Dicing** - dicing
* **?** - utforske tallmaterialet på flere aggregeringsnivåer
* **?** - SQL-konstruksjoner
* **?** - krysstabell
* **?** - standard gruppering og mengdefunksjoner
* **Subtotal** - subtotal
* **?** - langs n dimensjoner
* **?** - spørreresultatet
* **?** - nøkkelordene
* **?** - radene med nullmerker
* **Grand Total** - grand total
* **?** - visualiseringsverktøy

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

* **?** - trendanalyse
* **?** - konkurranseanalyse
* **?** - kundeanalyse
* **?** - markedsanalyse
* **?** - beslutningsstøtte
* **?** - vasking/rensking av data
* **?** - granulariteten av data
* **?** - nedflating av et hierarki
* **?** - stjerneflak (blanding av stjerne og snøflak)
* **?** - ekstraksjon
* **?** - rensing
* **?** - transformasjon
* **?** - summering
* **?** - lasting
* **?** - ad-hoc-spørringer
* **?** - spørreparametre
* **?** - analyse
* **?** - produksjonssystem
* **?** - stjerneskjema
* **?** - virksomhetskritiske data
* **?** - opprette
* **?** - slette
* **?** - endre
* **Transaction** - transaksjon
* **ER-model** - ER-modell, entitetsrelasjons-modell?
* **Entity** - entitet
* **Attribute** - attributt
* **relation** - relasjon
* **?** - relasjonsrolle
* **Aggregate Functions** - matematiske funksjoner som brukes for å telle antall poster i en relasjon
* **Atomic Value** - atomisk - udelelig verdi
* **Attribute** - attributt, egenskap
* **Candidate Key** - kandidatnøkkel
* **Cardinality Ratio** - kardinalitet
* **Cardinality Product** - kartesisk produkt
* **Cascading deletion** - kjedereaksjon
* **Complete Set** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc
* **?** - xxxxxxxxxxxxxxxxc

Ingen ordliste er komplett! Finner du ikke ordet du søker etter i ordlisten vår kan du forsøke språkrådets liste over [avløserord](http://www.sprakradet.no/Sprakhjelp/Raad/Norsk-for-engelsk/Avloeysarord/) og [termwiki](http://www.termwiki.sprakradet.no./wiki/Kategori:Språkrådets_datatermgruppe) samt universitetets fellesordliste for [dataord](http://i18n.skulelinux.no/nb/Fellesordl.eng-no.html).

Vi har også med vekslende hell benyttet [Google Translate](http://translate.google.com/#en/no/), [Babylon Translation](http://translation.babylon.com/english/to-norwegian/), [Wikipedia](http://no.wikipedia.org/wiki/Business_Intelligence)og [Store norske leksikon](http://snl.no). Oversettelsesverktøyene på nettet kommer ofte opp med merkelige svar som f.eks. "korn-erklæringen" for "Grain Declaration", mens de norske oppslagsverkene ofte mangler definisjonen på ordene.

Du kan også prøve [terminologi databasen til SAP](http://sapterm.com/).

Under finner du en liste over tekster som omhandler datavarehus:

* [Databasesystemer - Datavarehus (Universitetsforlaget)](http://www.dbsys.info/Databasesystemer/1_Ekstrastoff/datavarehus.pdf)
* [Databaseadministrasjon og -systemer - Datavarehus (Høgskolen i Østfold)](http://www.it.hiof.no/das/Datavarehus_2011.pdf)
* [Business Intelligence uten et tradisjonelt datavarehus (Universitetet i Agder)](http://brage.bibsys.no/xmlui/bitstream/id/85936/IS-501%202013%20vår%20masteroppgave%20Kristoffer%20Wang.pdf)
* [Opplæring i Business Intelligence - Introduksjon av Datavarehussenter (IBM)](ftp://public.dhe.ibm.com/ps/products/db2/info/vr8/pdf/letter/nlv/db2tun80.pdf)
* [Opplæring i Business Intelligence - Utvidede leksjoner i datavarehus (IBM)](ftp://public.dhe.ibm.com/ps/products/db2/info/vr8/pdf/letter/nlv/db2tan80.pdf)
* [Excel grunnleggede - kompendium (Tore Søfting)](http://www.excelkurs.com)
* [Exel avansert - kompendium (Tore Søfting)](http://www.excelkurs.com)
* [IBM InfoSphere Data Architect - Dimensional Modeling (IBM)](http://pic.dhe.ibm.com/infocenter/idm/docv3/index.jsp?topic=%2Fcom.ibm.datatools.dimensional.ui.doc%2Ftopics%2Fc_dm_fact_tables.html)
* [Datamodellering - tips og oppgaver (Trell)](http://www.trell.org/ikt/datamod.html)
* [Modeller for design av Web-Applikasjoner (UiB)](http://www.ii.uib.no/~khalid/INF329-H04/INF329-kursmaterial/inf329-h04-Kap02.pdf)
* [Web-teknologier og dataimplementasjon (UiB)](http://www.ii.uib.no/~khalid/INF329-H04/INF329-kursmaterial/inf329-h04-Kap11.pdf)
* [Datamodellering ordliste (UiS)](http://www.ux.uis.no/~aurlijo/ordliste.htm)