17. januar 2014

**Praktiske erfaringer med SelfServiceBI**

[1. SelfServiceBI på Videncentret 1](#_Toc376626379)

[1.1 Tekniske forudsætninger 2](#_Toc376626380)

[1.2 Etablering af datawarehouse på SelfServiceBI server 3](#_Toc376626381)

[1.3 Adgang til datamodeller 4](#_Toc376626382)

[2. Modelbygning med SelfServiceBI og ”Power-udvidelsespakkerne” til Excel 5](#_Toc376626383)

[2.1 Powerpivot 5](#_Toc376626384)

[2.2 Tidsserier 5](#_Toc376626385)

[2.3 Powerview 6](#_Toc376626386)

[2.4 Fra analyse til præsentation 6](#_Toc376626387)

[2.5 Uddannelsesbehov 6](#_Toc376626388)

[2.6 Før vi kan komme i gang 7](#_Toc376626389)

[3. Brugsfasen 7](#_Toc376626390)

## 1. SelfServiceBI på Videncentret for Landbrug

I efteråret 2013 er der gennemført et pilotprojekt på Videncentret for Landbrug for at afprøve potentialet i SelfServiceBI i forhold til VFL´s databaser med landbrugsfagligt indhold. Begrebet SelfServiceBI beskriver et IT setup, der gør det muligt for medarbejdere i en virksomhed, at få adgang til grunddata og frit lave udtræk uden at det er nødvendig at involvere IT afdelingens databaseeksperter.

Årets forretning 2013 med titlen ”Fra MIPS til forandring” er et godt eksempel på det forretningsmæssige potentiale i SelfServiceBI. Forretningen går ud på at sammenstille data fra forskellige datakilder og udlede viden, der kan bruges til at ændre på arbejdsmæssige rutiner hos DLBR medarbejdere. Arbejdet med data i ”Fra Mips til forandring” er baseret på en række manuelle rutiner, men det kunne let rationaliseres ved hjælp af SelfServiceBI.

Økonomiteamet i VFL-IT har i foråret 2013 anvendt SelfServiceBI i praksis til at holde øje med den daglige udvikling i Økonomidatabasen på en tv skærm. Det har eksempelvis været muligt at følge udviklingen i antal overførte ejendomme fra Ø90 til Økonomidatabasen, og at holde øje med fejlraten i de overførsler der er foretaget. Det er oplysninger, som tidligere krævede en del arbejde med SQL direkte i Økonomidatabasen, mens det nu sker helt af sig selv, hver gang modellen opdateres med nye data fra produktion.

**Sammendrag**

SelfServiceBi projektet har vist at det med lille indsats kan åbne for adgang til VFL´s data via Excel, og at Videncentret for Landbrug næsten har det nødvendige hard- og software til at gå i gang med SelfserviceBI. Barrieren for at komme i gang med SelfserviceBI på Videncentret for Landbrug er derfor ret lav, men det kræver at nogen har kendskab til mulighederne, og ser et forretningsmæssigt potentiale i SelfServiceBI.

I projektet er der arbejdet med data fra Økonomi- og Kvægdatabasen, samt fra IIS logfiler fra webservere. Projektet har vist, at data skal stilles til rådighed for SelfServiceBI via et datawarehouse, og at den del skal udvikles specielt til formålet.

Når datawarehouset er klar er der næsten ubegrænsede anvendelsesmuligheder i Excel. Hele Excel værktøjskassen kan nemlig tages i brug på de indlæste data. Projektet har vist at det er vigtig at afgrænse data i forhold til analysens formål, fordi for store datamængder kan resultere i hastighedsproblemer i Excel.

Projektet har også vist udfordringer i forhold til kvaliteten af de data, der indgår i en analyse. Fri adgang til data stiller store krav til analysemedarbejdernes datakendskab, fordi de ellers kan komme til at lave fejlagtige analyser eller fejlfortolke analyseresultaterne. Det sammenholdt med særlige krav til Excel kompetencer taler for at medarbejdere, der skal arbejde med SelfServiceBI skal være uddannelsesmæssigt godt klædt på til opgaven.

Sammenfattende kan det konkluderes, at der er store muligheder og potentiale i at arbejde videre med SelfServiceBI på Videncentret for Landbrug.

## 1.1 Tekniske forudsætninger

De tekniske muligheder for at lave SelfServiceBI ved hjælp af Microsoft produkter har været under kraftig udbygning i de seneste år, fordi der er udviklet nye datalagringsteknologier. Kombinationen mellem Microsoft Office og enterprise udgaverne af SQL Server 2012 eller SQL Server 2014 giver et godt grundlag for SelfServiceBI.

I SelfServiceBI projektet er de nye tekniske muligheder kombineret i et setup, der i skematisk form ser sådan ud:



Tegningen illustrerer opdelingen mellem kildesystemet med grunddata på VFL, en særlig SelfServiceBI server, der stiller bearbejdede data til rådighed og selve den arbejdsplads, hvorfra data kan udnyttes.

Det viste på tegningen kan etableres i et driftssetup på Videncentret for Landbrug uden store investeringer, men det vil dog være nødvendig at upgradere mindst en SQL Server fra SQL Server 2008R2 til enterprise udgaven af SQL Server 2012 eller 2014.

Det vil endvidere være en fordel, at de medarbejdere der skal arbejde med SelfServiceBI har en pc med meget ram og Office 2013 installeret. Det er dog ikke en nødvendig forudsætning (se kom godt i gang med SelfServiceBI).

Det tekniske setup i SelfServiceBI projektet har bestået af en kraftig 64bit PC med i7 processor og 20GB ram. PC´en har fungeret som SelfServiceBI server med Windows Server 2012 som styresystem. Serveren har været indmeldt i LC.SKEJBY domænet, og de brugere der har haft adgang til serveren er oprettet som brugere på serveren, med de nødvendige rettigheder. Der er installeret følgende på serveren:

* SQL server 2012 enterprise edition incl. Analysis Services, Integration Services og Reporting Services
* Microsoft SQL Server Data Tools
* Business Intelligense for Visual Studio 2012

Det har været muligt at etablere setuppet udelukkende baseret på MSDN udviklerlicenser og der er derfor ikke foretaget andet indkøb end PC´en.

## 1.2 Etablering af datawarehouse på SelfServiceBI server

Når hardware og nødvendige programmer er på plads er næste skridt at få data gjort tilgængelige i en form, så de er egnede som input til Excel. En tabel er egnet som input til Power Pivot, hvis den har en forståelig kolonneopdeling.

I SelfServiceBI projektet er der arbejdet med grunddata fra tre datakilder:

1. Økonomidatabasen
2. IIS webserverlog
3. Kvægdatabasen

Det har for alle datakilder vist sig nødvendig at bearbejde data før de var egnede som grundlag for SelfServiceBI arbejdet.

For Økonomidatabasen skyldes det primært, at data er normaliseret, men også at mange faglige oplysninger er gemt på referencenumre, som skal oversættes til kolonneoverskrifter, før det er muligt at arbejde effektivt med dem i Excel.

IIS webserverloggen er en flad tekstfil, som skal oversættes til en regulær kolonneopbygning før den er egnet til SelfServiceBi.

I SelfServiceBI projektet har der ikke været direkte adgang til Kvægdatabasen, men kun adgang til et udtræk fra databasen i en flad fil. Det kan derfor ikke umiddelbart vurderes, hvor egnede data i Kvægdatabasen er til SelfServiceBI uden bearbejdning, men der er grund til at tro, at udfordringerne vil minde om dem der har været med data fra Økonomidatabasen.

Som illustreret på tegningen ovenfor består bearbejdningen af data på SelfServiceBI serveren af tre komponenter:

1. ETL
2. Datawarehouse
3. Kube / Tabular

ETL delen (extract, transform, load) går ud på at få udtrukket de relevante data og få dem sammenstillet på en måde, så de kan stilles til rådighed som tabeller eller views i datawarehouset. Load delen drejer sig om at få automatiseret processen med at hente de seneste data fra grunddata, så de data der stilles til rådighed via datawarehouset er ajourført i den ønskede takt.

Datawarehouset udstiller de data det er muligt at arbejde med i Excel enten i form af tabeller eller views.

Kube /Tabular delen indeholder Excel bearbejdede datamodeller. Når en datamodel er gemt på SelServiceBI Serveren, kan den genbruges som grundlag for nye eller flere forskellige dataanalyser i Excel, og det er dermed muligt at genbruge alt det arbejde, der er lavet i Excel for at udlede det fagligt relevante datagrundlag.

I forbindelse med SelfServiceBI projektet har det været nødvendig at løse følgende opgaver for at få etableret SelfServiceBI serveren.

Der er lavet et program, der kan oversætte referencenumre til kolonneoverskrifter i forbindelse med restore af SelfServiceBI databasen. Det viste sig, at der findes mere end 3.500 referencenumre i Økonomidatabasen og at transformationen fra referencenumre til kolonner gav en tabel med mere end 3.500 kolonner. Det var et problem, fordi der findes en begrænsning i MSSQL server, der gør at en tabel max kan have 1024 kolonner og max 8kb i en række.

Opsplitningsprogrammet blev derfor udvidet med en funktion, der kan opdele i tabeller ud fra type og med en kolonnebegrænsning på max 500 kolonner pr. tabel.

Arbejdet med data i Excel viste, at det også var nødvendig at lave nye ID´er til nogle af tabellerne, og der blev lavet en funktion, der kan autogenerere de nødvendige ID´er i forbindelse med den natlige kopiering af databaser fra produktion. Behovet for nye ID´er vil højst sandsynlig opstå som følge af den lette adgang til data via Excel, som vil give nye idéer til hvilke analyser der kan laves.

Der er endvidere lavet et program, der kan transformere IIS logfiler til en databasetabel med relevante kolonner og oplysninger, og udtrækket fra Kvægdatabasen er tilsvarende transformeret til tabeller i SQL databasen.

Alt ovenstående er automatiseret i et natlig job, som også restorerer databaserne i SelfServiceBI datawarehouset ud fra backupfiler fra Økonomidatabasen produktion.

## 1.3 Adgang til datamodeller

SelfServiceBI modeller som er udarbejdet i Excel kan gøres tilgængelig for andre analytikere ved at lægge modellen i form af en tabular kube på en Analysis Server. Serveren kan sættes til automatisk at opdatere modellen, så den f.eks hver nat bliver opdateret med de seneste data fra SelfServiceBI datawarehouset. Derved har andre analytikere altid adgang til de seneste data i modellen, og man skal som analytiker ikke selv huske at opdatere sit datagrundlag.

En model der stilles til rådighed for en bruger svarer til at give adgang til et regneark med data. I forhold til datasikkerhed skal en datamodel altså håndteres på samme måde, som et almindeligt regneark, fordi regnearket indeholder alle data og dermed kun skal være tilgængelig for dem der må se data. Adgangen til modellen kan styres via roller i Analysis Server, mens Excel modellens data er tilgængelige, for alle der har adgang til Excel filen. Det gør altså ikke meget nytte, med adgangsstyring i Analysis Server, hvis regnearksfilen bliver placeret på et fællesdrev, hvor alle kan få fat i den.

## 2. Modelbygning med SelfServiceBI og ”Power-udvidelsespakkerne” til Excel

Et af formålene med at gennemføre projektet har været at undersøge hvilke muligheder, man som analytiker får med SelfServiceBI i forhold til de eksisterende værktøjer, som vi anvender til udtræk og analyse af data fra Økonomidatabasen (og andre databaser).

## ****2.1 Powerpivot****

Indledningsvis blev det besluttet ved hjælp af Powerpivot at arbejde på at genskabe det datamateriale, som var grundlaget for udarbejdelsen af hæftet ”Business Check Slagtekalve”. I sin oprindelige form er dette datasæt trukket ud med ”ProdAna”, som er et udtræksværktøj, der primært anvendes til at lave dataudtræk for driftsgrensanalyser fra Ø90. Det oprindelige datasæt er, udover data fra Ø90, suppleret med oplysninger om årsdyr (foderdage), som er hentet fra Kvægdatabasen.

Via Powerpivot kan man få adgang til data direkte fra forskellige databaser. Det er således ikke bearbejdede data, som vi kender det fra f.eks. KredsAna og ProdAna. Fordelen ved det er, at man ikke er bundet af et bestemt udvalg af variable. Det vil sige, at det er muligt alene at hente de (få) variable, man skal bruge til den givne analyse, og da det er nemt at koble flere variable på, hvis man opdager et behov for det undervejs i analysen af data, er der ingen problemer i, at man begrænser sig i opstarten af analysen. Det er desuden muligt at koble data fra flere databaser sammen i Powerpivot. I eksemplet med Business Check Slagtekalve er det data fra Økonomidatabasen og Kvægdatabasen. Sammenkoblingen kræver, at der er en fælles ident. Det kan f.eks. være cvr-nr eller lign. Endelig er det muligt ud fra de udvalgte variable at beregne de ting, man ønsker, og bruge det i analysen på lige vilkår med de oprindelige data.

Når man har adgang direkte til data, stiller det større krav til analytikerens kendskab til data. I dag måler vi f.eks. på en række parametre for at afgøre, om et regnskab og/eller driftsgrensanalyse er egnet til at indgå i et datasæt. Det skal analytikeren selv være opmærksom på, når han får direkte adgang til data.

En vigtig forudsætning for, at det giver mening at arbejde med Business Check og andet på driftsgrensanalyser eller lign. er, at datamodellen bliver placeret på en central server, som alle analytikere har adgang til. Det er nødvendigt, da arbejdet med at udarbejde datamodellen fra gang til gang er for omfattende. Ydermere vil hver analytiker udarbejde meget enslydende datamodeller, hvilket også taler for at have én central datamodel, som hver analytiker kan tilpasse til den givne analyses specifikke formål.

Arbejdet med data i Powerpivot har vist sig – i nogen grad – at være intuitivt. Det er en udfordring, at værdierne fra de underliggende tabeller i databasen skal hentes og beregnes vha. DAX. Med få og enkle instruktioner er det dog forløbet ganske fornuftigt, og ikke mindst autocompletefunktioner på variabelnavne har gjort arbejdet med at kode specifikationen væsentligt mere sikker i forhold til fejlindtastninger. Det gælder især sammenlignet med det tilsvarende arbejde, der løses i OEDB-Admin, når specifikationen til ProdAna skal tilpasses.

Udover indbygget autocomplete er det en stor hjælp i modelleringsfasen, at alle tallene er tilgængelige, og at enhver beregning dermed udføres med det samme. På den vis valideres formlernes indhold ved at se på, om det er rimelige værdier, der kommer ud.

## ****2.2 Tidsserier****

Det er også undersøgt, om vi kan anvende Powerpivot til at analysere paneldata. I regi af Økonomidatabasen betyder det, at vi kan udtrække regnskaber for de samme ejendomme over længere tid end to år. Perioden på to år er en begrænsning i vores nuværende udtræksværktøjer – KredsAna og ProdAna.

Selv om der indtil videre kun er anvendt kort tid på at undersøge mulighederne for at trække data over tid ud med Powerpivot, vurderer vi, at der er et potentiale i at lave denne type analyser via Powerpivot. Det kræver dog formentligt, at man i forbindelse med opbygning af datamodellen er meget omhyggelig med ikke at vælge for mange variable ud. Med den gode mulighed, der er for at opdatere modellen med nye variable undervejs i analysearbejdet, er det ikke et problem.

## ****2.3 Powerview****

Man kan få stor nytte af Powerview undervejs i modelleringsfasen. Powerview er en anden udvidelsespakke til Excel. Ved hjælp af Powerview kan der nemt oprettes grafiske visninger i flere typer, som tilpasser sig de filtre og variable, som udvælges til visning. Det gør det muligt for analytikeren at danne sig et hurtigt overblik over data uden at skulle trække samtlige variable ud i en given sortering. Det forbedrer muligheden for at fordybe sig i den egentlige analyse, inden data bliver stillet op til præsentation.

## ****2.4 Fra analyse til præsentation****

Powerpivot sørger for at samle data og gør brug af relationer mellem de enkelte tabeller. Dette sker i ”administrér-billedet”, mens data som udgangspunkt bliver præsenteret i noget, der til forveksling ligner en almindelig pivottabel. Dermed følger udfordringen i, at så længe data er ”levende”, dvs. så længe, der er en forbindelse ned til databasen, og man dermed har mulighed for at ændre i såvel variable som filtre og sorteringer, er det ikke umiddelbart muligt at lave en præsentation af hele tabellen på den vis, som det sker i hæftet Business Check Slagtekalve. Tabellerne fra Business Check Slagtekalve er lavet i almindeligt Excel-format med de formateringsmuligheder, der følger med der.

Såfremt man ønsker at lave sådanne præsentationer, er det vores opfattelse, at det skal ske ved, at man kopierer værdierne fra pivottabellen til et ark, hvor man efterfølgende kan lave en egentlig præsentation. Powerpivot er således alene et analyseværktøj og ikke et præsentations- eller rapporteringsværktøj.

## ****2.5 Uddannelsesbehov****

For at vi kan udnytte det fulde potentiale i Powerpivot, Powerview osv. er det nødvendigt, at de analytikere, der skal arbejde med værktøjerne, kommer på kursus i det. Det skal gerne være kurser, hvor der er mulighed for at arbejde med egne data undervejs.

Derudover skal vi være opmærksomme på, at Powerpivot stiller andre (og formentlig større) krav til os som analytikere, hvilket betyder, at det ikke er nok med værktøjskendskab, men at vores tilgang til analyser også skal ændres. Herunder ligger også de kvalitetsparametre, som i dag er trukket ned over vores udtræksprogrammer, men som vi i f.eks. Powerpivot selv skal have styr på. I de nuværende udtræksløsninger er der også ”basismodeller” som genbruges fra gang til gang, og dermed sørger for at medtage kriterier for kvalitetskontrol. Disse kriterier kan som sådan sagtens blive genbrugt i arbejdet med Powerpivot. Den risiko, der beskrives ovenfor, er, at analytikeren slet og ret glemmer at medtage disse kriterier.

Arbejdet med dataanalyse i Powerpivot betyder, at analytikeren får ”levende” data i hånden. Dette giver muligheden for at arbejde i modellen, mens den bliver til. Dette er markant anderledes end udtræk med KredsAna, hvor KredsAna på sin vis også er en rapportgenerator. Analyse-tiden i arbejdet med KredsAna foregår i grove træk i den tid, man bruger på at vælge kriterier, gruppering og sortering, hvorefter data kommer ud og bliver statiske.

## 2.6 Før vi kan komme i gang

Noget, der kan gøre analysearbejdet mere enkelt, er, at man som analytiker får adgang til beregnede felter, der følger den kendte navngivning (opsætning) fra et KredsAna-udtræk – dvs. en opstilling med linjer, der stort set svarer til et landbrugsregnskab. Det gør det muligt hurtigt og overskueligt at tilvælge de linjer (felter/søjler), som man i første omgang interesserer sig for, og hurtigt gøre data levende vha. Powerview. Det kan eksempelvis være en sammenhæng mellem udsæd, gødning og bruttoudbytteplanteavl, hvor modeludvidelsen består af at inddrage data fra Dansk Markdatabase vedr. markstørrelse eller form. Som data findes i dag, er man som analytiker hensat til at bygge modellen fra bunden, dvs. fra de referencenumre, som Økonomidatabasen er baseret på. Det betyder, at man starter med ref.nr. 000100 (korn) og fortsætter med 000230 udsæd, alt imens man vælger at undlade 000250 korn, da det er en omkostningspost til foderkorn, og samtidig undlader 021000 mfl., som er andre referencenumre, der indeholder oplysninger om udsæd.

Det forventes, at en sådan opbygning af ”basislinjer” til fælles brug kan blive en forholdsvis stor opgave, såfremt den nuværende KredsAna-specifikation ikke kan udlæses/genbruges til formålet.

## 3. Brugsfasen

I SelfServiceBI projektet er vi ikke nået til en reel brugsfase af de opbyggede modeller, fordi al den afsatte tid er gået til klargøring af datawarehouse og udvikling af modeller i Excel.

Som nævnt i indledning har økonomiteamet i VFL-IT anvendt SelfServiceBI i foråret 2013 til at holde øje med den daglige udvikling i Økonomidatabasen på en tv skærm. Det har været muligt at følge udviklingen i antal overførte ejendomme fra Ø90 til Økonomidatabasen og at holde øje med fejlraten i de overførsler der er foretaget. Det er oplysninger som tidligere krævede en del arbejde med SQL direkte i Økonomidatabasen, mens det nu sker helt af sig selv hver gang modellen opdateres med nye data fra produktion.

Potentialet i SelfServiceBI må kunne bruges i mange sammenhænge. Eksempelvis kan det udnyttes til et dashboard, der viser det økonomiske resultat for forskellige driftstyper efterhånden som regnskaberne bliver overført til Økonomidatabasen, eller til regional analyse og visning af resultater og tendenser præsenteret på et landkort.