



L&F / SEGES Analytics og BI-værktøjer

Anbefaling af analyseværktøjer

30/11

Zangenberg har udvalgt og evalueret en række BI-platforme som matcher de behov, vi ser i fagdomænerne i L&F / SEGES, og som kan danne grundlag for mere datadreven rådgivning

- Et nyt værktøjsvalg eller anderledes organisering af håndtering af data skal være bundet i eksisterende behov og ønsker fra fagområderne. For at sikre det bedste værktøjsvalg, har Zangenberg foretaget en evaluering af behov og faglige ønsker i de forskellige fagdomæner.
- Evalueringen har taget afsæt i:
 - i. de udfordringer, der i dag løses under de enkelte fagdomæner. Herunder også inkludering af nuværende BI-værktøjer og analysearbejde
 - ii. hvilke nye opgaver og løsninger, der tænkes at kunne integreres indenfor de næste 2-3 år. Sidstnævnte særligt set i lyset af den data (type, kvalitet, mængde) der tænkes at blive tilgængelig i den nærmeste fremtid.
- På baggrund af behovsafdækningen har vi udvalgt og evalueret en række BI-platforme som matcher de behov vi ser i fagdomænerne, og som kan danne grundlag for mere datadreven rådgivning. Evalueringen er blandt andet set i lyset af, hvilke værktøjer der allerede eksisterer rundt i fagdomænerne **24-36 md.** Arbejdet med avanceret analyse er meget forskelligt imellem domæner, og flere steder bliver værktøjet dikteret af data- og opgavetyper. Behovene er dog mere ensrettede for mere basal dataanalyse og –visualisering. Her vil der være potentialer i, at Koncern IT implementerede et BI-værktøj på tværs af domæner, der således kan benytte egne, mere avancerede analyseværktøjer på toppen. Både QlikView, Tableau, SAS, Cognos og Power BI er værktøjer, der kan løfte denne opgave. Evalueret ud fra funktionalitet, pris, supportmodel og integration med eksisterende værktøjer i L&F/SEGES, finder vi at Microsoft Power BI vil være det mest oplagte værktøj at implementere på tværs.

Evaluering af BI-værktøjer

Vi har evalueret en række værktøjer baseret på bla. funktionalitet, pris og integration til eksisterende værktøjer i L&F / SEGES

- Vi har analyseret en række værktøjer med udgangspunkt i de eksisterende platforme og behov i L&F / SEGES. Foruden inddragelse af markedsledende leverandører, har vi valgt værktøjerne ud fra hvad der allerede findes i domænerne og input fra samtalerne.
- Valget af BI-platforme skal være bundet af de eksisterende og fremtidige behov i fagområderne. Et af hovedkriterierne for evaluering af BI-platforme er derfor **funktionalitet**. Værktøjerne er evalueret efter hvordan de opfylder både nuværende og fremtidigt ønskede behov, eksempelvis behovet omkring bedre udstilling og visualisering af data.
- Dernæst har man en lang række eksisterende værktøjer, kompetencer og specialløsninger i L&F / SEGES. Et andet hovedkriterie for evaluering er derfor **integrering med eksisterende værktøjer og datakilder**. Vælger man eksempelvis en ny BI-platform til descriptive analytics, er det essentielt at værktøjet både kan håndtere de benyttede datakilder og integrere med platforme brugt til mere avanceret analytics.
- Foruden hovedkriterierne er alle værktøjer evalueret i forhold til andre, væsentlige kriterier som **pris, brugervenlighed og dansk supportnetværk**.

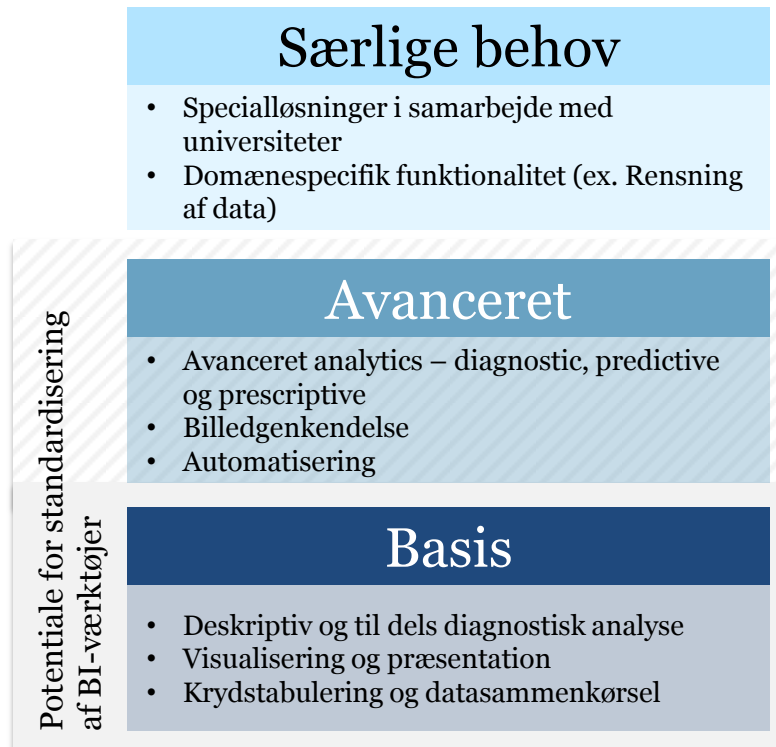
Analyserede værktøjer



Note: Vi har derudover vurderet værktøjer fra SAP, Elastic og KNIME

Der er potentialer for standardisering af BI-værktøjer for at dække behov, der går på tværs af domæner

- Zangenberg ser potentialer i, at Koncern IT sætter linjen for at løse behov, der går på tværs af koncernen. Vi har opdelt behov i 3 kategorier:
- **Basal**
 - De behov, der knytter sig til deskriptiv og diagnostisk analyse kan løses med klassiske BI-værktøjer
- **Avanceret**
 - De behov, der bla. knytter sig mere avanceret analyse og kræver funktionalitet udover klassiske BI-værktøjer
- **Særlige behov**
 - Domænespecifikke behov, der enten knytter sig til specifikke projekter, arbejdsområder eller bestemmes ud fra eksogene faktorer (som samarbejde med universiteter)



Eksempler på værktøj

- Løsninger der løbende udvikles

- R
- SAS
- Python
- Tensorflow

- SAS VA
- Cognos
- PowerBI
- Tableau
- QlikView

Størstedelen af domænerne ønsker et bedre værktøj til descriptive analytics og datavisualisering samt en interaktiv "glasplade" til landmanden / rådgiveren

Basisbehov

- Basisbehov dækker over udstilling af data, visualisering og generel deskriptiv (og diagnostic) analyse.
- Både QlikView og Tableau leverer meget god datavisualisering og integrerer godt med R (kræver dog udvidelse af QlikView). Tableau mangler dog blandt andet en række Enterprise features (som at integrere til Linux og printe til PDF og PP).
- L&F / SEGES har i øjeblikket SAS VA integreret flere steder, og bruger SAS til både diagnostic og predictive analytics. SAS integrerer dog mindre godt med tredjepartsapplikationer.
- IBM Cognos og MS PowerBI har god performance ift. L&F / SEGES basisbehov. Begge platforme tilbyder derudover en række løsninger til mere avanceret analytics (hhv. Cortana Intelligence og Watson Analytics). IBM Cognos integrerer dog ikke til R, og kun sparsomt til Excel.

	QlikView	Tableau	IBM Cognos	Microsoft PowerBI	SAS (EG, VA)	
Dækning af basisbehov	Platform til udstilling af data og resultater					
	Bedre visualisering af data					
	Værktøj til deskriptiv analyse					
Resterende kriterier	Integrering med eksisterende værktøjer				*	
	Brugervenlighed					
	Pris			**		
	Leverandør-ramme					

*I forvejen mange SAS løsninger i L&F / SEGES. Dog integrerer SAS ikke så godt med andre tredjepartsapplikationer

** Cognos tilbyder ingen standardpriser. Dog tyder intet på at løsningen er billigere end QlikView og Tableau

QlikView tilbyder god datavisualisering og dashboards. Ved en udvidelse kan man integrere til R og Python, men licensmodellen kan være omkostningsfuld



Beskrivelse

QlikView er et BI-værktøj til analyse og visualisering af data. Platformen indeholder blandt andet også rapportplanlægnings og -distributionsprogrammet Nprinting.

Gartner placerer Qlik (der både indeholder QlikView og Qlik Sense) under markedsledere, baseret på leverandørens indhold, scope, markedsandel og profitabilitet. Placeringen er grundet en robust og over-middel kundeoplevelse og et stærkt netværk af samarbejdspartnere.

Med QlikSense og Qlik Advanced Analytics Integration (AAI) kan brugere integrere med værktøjer som R og Python.

Styrker

- Nem og hurtig implementering
- Intuitivt, brugervenligt og visuelt appellerende
- Understøtter en lang række datakilder

Svagheder

- Færre muligheder for sofistikeret BI-arbejde og dataopdagelse. Integration til R og Python kræver licens af både QlikSense og Qlik AAI. Udvidelsen er derudover meget ny (Juni 2017) og modenheten er uvis
- Teknisk support ligger lidt under middel ift. ekspertise, responstid og løsningsstid
- Licensmodellen kan gøre QlikView til en dyr løsning. Særligt hvis man ønsker integration til R og Python

Tableau er guldstandarden i datavisualisering, men kan kun tilbyde væsentlige enterprise-features igennem tredjepartsaftaler



Beskrivelse

Tableau er af Gartner regnet for guldstandarden for visualisering og dataopdagelse. Foruden takling af beskrivende BI-opgaver, kan Tableau håndtere både store og heterogene datakilder.

I Tableau-pakken får brugere derudover i stigende grad mulighed for smart data discovery, Natural Language Processing, Maskinlæring, forecasting, geo-kodning og automatisering af rutineopgaver.

Dog mangler Tableau en række Enterprise-features som integrering med Linux og gode API'er og print til PDF og PowerPoint. Dette kan løses med eksterne programmer, som dog tilføjer ekstra omkostninger til en i forvejen dyr løsning.

Skønt Tableau er benyttet af mange, ligger supporten hovedsageligt i små virksomheder (som Invisio).

Styrker

- Ledende i datavisualisering og dataopdagelse
- Mulighed for automatisering og geo-kodning
- Kan håndtere større datamængder

Svagheder

- Mangel på Enterprise-features (print til PDF og Power-Point, Linux support etc.). Dette kan dog opnås via eksterne løsninger
- Mangel på support for komplekse datamodeller der integrerer flere datakilder på samme tid
- Omkostningsfuld

IBM's Cognos er brugervenlig og tilbyder mange muligheder indenfor basal dataanalyse. Platformen integrerer godt med resten af IBM, men mangler tredjepartsintegration



Beskrivelse

Cognos er IBM's BI-løsning. Platformen er brugervenlig og tilbyder en bred portefølje af muligheder indenfor basal dataanalyse og god visualisering af data.

Til mere avanceret analytics tilbyder Cognos en god integration med de andre værktøjer i IBM's analytics-portefølje, som eksempelvis Watson Intelligence. Cognos mangler dog integration til tredjepartsapplikationer, og integration til eksempelvis Python sker kun begrænset igennem deres SDK.

Cognos supporterer en lang række datakilder, men kan have problemer med at behandle store datasæt og håndtere ustruktureret data.

Styrker

- Interaktiv visuel dataopdagelse og dashboard
- Bred portefølje af muligheder indenfor basal analyse
- Gode muligheder for avanceret analyse igennem IBM's andre værktøjer
- Stort supportnetværk igennem IBM

Svagheder

- Mangler integration til tredjepartsapplikationer, der dog i flere tilfælde kan købes via produktudvidelser
- Kan have performance-problemer (ingen in-memory engine)
- Problemer med at håndtere ustruktureret data og store datasæt (eneste platform der ikke har en in-memory engine)

Microsofts PowerBI tilbyder en BI-plattform med mange muligheder for basal dataanalyse og god integration med Excel og R. Dog usikker skalérbarhed for storage og processing



Beskrivelse

Microsoft's PowerBI tilbyder flere muligheder indenfor basal dataanalyse, herunder dataopdatelse, datavisualisering og interaktive dashboards. Særligt er værktøjet rettet mod parameteriserede rapporter og dashboards.

Værktøjet indeholder i sig selv ikke mange muligheder for avanceret dataanalyse. Dog integrerer PowerBI godt med flere værktøjer som Excel, R og Cortana Intelligence, MS avancerede analyseplatform. Blandt andet understøtter PowerBI direkte eksekvering af R scripts. Integrationen af PowerBI med Python skal i øjeblikket ske igennem Microsoft SQL Server. Direkte kan Python bruges til at rengøre data, som kan streames ind i PowerBI.

Styrker

- God integration med R og Excel
- Integration til Python igennem MS SQL Server
- Smart data discovery
- Mulighed for udvidelser igennem Microsoft ML server og Microsoft R Open
- Integrering med dashboards og rapporter igennem API
- Pris
- Godt dansk support-community igennem MS

Svagheder

- Mangel på avanceret funktionalitet, specifikke formateringsmuligheder og support til Linux
- Usikker skalérbarhed for data storage og processing

SAS er allerede brugt i vid ustrækning i L&F / SEGES. Værktøjet tilbyder kompleks dataanalyse, men man oplever performanceproblemer og mangler tredjepartsintegration



Beskrivelse

SASs BI-løsning (Visual Analytics, Enterprise Guide og Office Analytics) giver en lang række analytiske muligheder, særligt ved brug af mere kompleks analyse og større datasæt.

SAS Visual Analytics tilbyder en interaktiv og visual dataopdagelse, mens de resterende platforme tilbyder mere kompleks dataanalyse.

Skønt SAS bruges i vid udstrækning i L&F / SEGES, integrerer det dog ikke godt med tredjepartsapplikationer. Derudover har enkelte domæner, som Svin – Datamanagement udtrykt utilfredshed med performance.

Styrker

- Kompleks dataanalyse
- Skalabilitet
- Visuel dataopdagelse

Svagheder

- Dårlig integration med tredjepartsapplikationer og dashboards. Integration til R og Python sker igennem udvidelser med hhv. SAS/IML og SAS Viya
- Utilfredshed med performance



BI-løsningerne tilknyttet store leverandører tilbyder et bredt supportnetværk, hvor support til løsninger fra mindre leverandører er mere fragmenteret








- Det er væsentligt, at den strategiske BI-partner skal have ressourcer til rådighed for supportering af løsningen – også udover implementeringsfasen. For en organisation som L&F / SEGES, der er dansk både af geografi og kultur, er det vigtigt at BI-leverandøren har et etableret dansk supportnetværk.
- Både Cognos, SAS og Power BI er løsninger udbudt af virksomheder, der alle har et stort dansk fodaftryk. For at være IBM eller Microsoft certificeret ekspert kræver det evner til at kunne supportere hhv. Cognos og Power BI. Det betyder, at disse værktøjer følger med en omfangsrig dansk ekspertise til at supportere løsningen. Det samme gælder SAS, der har et stort marked i Danmark.
- For værktøjer som Tableau og QlikView der, væksten til trods, stadig tilhører nicheprodukter, er supportnetværket mere smalt og fragmenteret. Flere mindre virksomheder og rådgivere, som ex. Invisio og Inspari, tilbyder hjælp med implementering og brug. Disse rådgivere fungerer dog mere som salgskontorer, og det kan være vanskeligt at finde support i tilfældet af problemer. Det mindre danske marked gør også, at rådgivere specialiserer sig i flere værktøjer på samme tid.
- Flere leverandører vil derudover have vanskeligt ved at håndtere problemer på tværs af softwarelag. I tilfælde af problemer på tværs af ex. BI-værktøj og Data Warehouse, vil man i mange tilfælde skulle kontakte forskellige leverandører. Værktøjer som Power BI, der er en del af MS stakken, giver i stedet mulighed for support i hele kæden.



Zangenberg anbefaler PowerBI til at dække de basale behov, der går på tværs af fagdomæner i L&F / SEGES

- Baseret på de basisbehov, vi har set i L&F / SEGES, samt det nuværende brug af analytics og BI-platforme, vurderer Zangenberg **Microsoft Power BI** til at dække størstedelen af de behov, der går på tværs af fagdomæner i L&F / SEGES.
- Foruden bred funktionalitet i deskriptiv BI-analyse, tilbyder Power BI gode muligheder for at tilgå data via en interaktiv platform (med adgangs- og indholdsstyring).
- Særligt er det Power BIs integration med MS-suiten og de eksisterende værktøjer i L&F / SEGES der gør værktøjet relevant. Dertil kommer muligheden for udvidelsen med eksempelvis Cortana for mere avanceret analytics (for behov der rækker udover R, som integrerer godt med Power BI). For alle værktøjer gælder det, at man i enkelte tilfælde kan være påkrævet forskellige produktudvidelser eller tredjepartsapplikationer for at opnå den ønskede funktionalitet. Vi vurderer MS Power BI til at være løsningen, der opfylder flest behov på én gang.
- Sidst er Power BI langt billigere end andre visualiseringsværktøjer som QlikView og Tableau, givet det kommunikerede behov.



Platform til udstilling af data og resultater	
Bedre visualisering af data	
Værktøj til deskriptiv analyse	
Integrering med eksisterende værktøjer	
Brugervenlighed	
Pris	
Leverandør-ramme	

Der er ikke stor potentiale i at udstikke en retning for værktøjer, til at løse avancerede behov. Derimod værdi i organisering af datasammenkørsel på tv

Avancerede behov

- De mere avancerede behov dækker over predictive og prescriptive analytics, billedgenkendelse og til dels automatisering af databehandling (skønt sidstnævnte kan gøres ved brug af værktøjer for basale behov).
- I fagdomænerne er de mere avancerede behov mere projektbestemt. I planteområdet koncentrerer avanceret analysearbejde eksempelvis til analyse af bearbejdet (og evt. ubearbejdet) GIS data. Billedgenkendelse efterspørges også i Svin – Datamanagement, i form af automatisk at kortlægge adfærdsmønstre baseret på overvågningsvideoer.
- I Kvæg – datamanagement søger man et værktøj, bedre end R og SAS, specifikt til automatisering af daglige eller ugentligt opgørelser via en HTML side, som brugere kan klikke sig ind på. Andre steder, som KØS, handler automatisering i stedet om nemt at kunne danne parametriserede rapporter – en øvelse der i øjeblikket gøres i LF Økonomisk Analyse og ellers kan tilgås via et basalt BI-værktøj.
- De opgavespecifikke behov betyder også, at fagdomænerne benytter en række forskellige værktøjer til avanceret analytics, hvor de største er R, SAS (og Python). Skønt enkelte domæner hverken er teknisk eller kulturelt bundet til SAS eller R, er flere løsninger stærkt bestemt ud fra opgavetyper (Macrobound og Mintel i LF – Analyse; specialudvidelser til SAS og R i Svin – Avl og Genetik; Tensorflow i PlanteInnovation etc.). Vi ser derfor ikke stor gevinst i at man udstikker en retning for værktøjer til at løse mere avancerede behov. Potentialet kan i stedet ligge i, at tilbyde god integration til basale værktøjer, som PowerBI, hvis man i fagdomænerne vælger at bruge R.
- Dog udtrykker alle fagdomæner et ønske om at sammenkøre data på tværs af domæner. Der kan derfor være gode samarbejds potentialer i, at man stiller centrale kompetencer til rådighed som kan administrere data på tværs af domæner.

Der skal etableres en organisering omkring sammenkørsel af landmandens stamdata

- Størstedelen af fagdomænerne udtrykker et ønske om at sammenkøre flere data på tværs af SEGES. Konkret er ønsker flere domæner at berige nuværende analyser med data fra andre områder. I forlængelse af et valg af BI-værktøj til basisbehov i SEGES, vil der være potentialer i at samle dele af den eksisterende data i et Datawarehouse som de forskellige domæner kan tilgå. Data kan samles centralt og administreres af kompetencer i Koncern IT.
- Fagdomænernes data kan overordnet inddeles i to kategorier: stamdata og ad hoc (forsøgs) data.
- **Forsøgsdata** er data der bliver samlet ind i forbindelse med forsøg eller ”ad hoc” analyser. Hvor nogle forsøgsdata er både meget detaljeret og af god kvalitet (som eksempelvis i KvægInnovation og Avl & Genetik), er andet data mindre detaljeret og af blandet kvalitet. Fælles for ”forsøgsdataen” er, at data kommer ad hoc, og der er ikke nødvendigvis systematik i hverken opsamling, lagring eller brug.
- **Stamdata** er data om landmanden, der bliver indsamlet systematisk i de eksisterende databaser. Det inkluderer detaljeret data omkring bla. økonomi, produktion og besætning, og ligger i databaser som Ø90, Markdatabasen og Kvægdata-basen. Fælles for ”stamdataen” er, at data kommer ind på en kontinuert basis og indeholder mange generelle informationer om landmanden.

Forsøgsdata

- Data brugt i forsøg, rullende afprøvninger, specialanalyser etc.
- Data er ad hoc og kan afvige i type og kvalitet fra gang til gang
- Meget specifikt ift. pågældende forsøg el. analyse

Stamdata

- ”Grunddata” om landmanden, herunder data omkring besætning, mark, produktion, økonomi etc.
- Data kommer på en kontinuert basis og fødes ind i eksisterende databaser (kvægdata-basen, markdatabasen etc.)

Der er potentialer i, at en delmængde af landmandens stamdata bliver administreret centralt af Koncern IT. Fagdomænerne kan internt koordinere omkring deling af resterende data

- For L&F / SEGES er der potentialer i at sammenkøre landmandens stamdata. Med sammenkørsel menes, at alle fagdomæner nemt og overskueligt skal have adgang til en delmængde af landmandens stamdata, uanset hvilket fagdomæne man tilhører.
- Der eksisterer i øjeblikket en stor mængde stamdata, og den stigende digitalisering af landbruget vil betyde en yderligere stigning i datamængder. Koncern IT skal tage en differentieret tilgang, og tage stilling til hvor stor en del af den eksisterende stamdata der skal administreres centralt. Den resterende del kan således administreres af de respektive fagdomæner og som kan tilgås af andre på opfordring.
- Bemærk, at der eksisterer stor variation af datatype, kvalitet og detaljeringsgrad indenfor stamdata. Mange af de udtrykte ønsker i fagdomænerne handler om at krydse landmandens data omkring produktion, bestand og markprofil, hvor andet, mere detaljeret data, kan være mindre relevant.
- En central administration af data vil kræve ressourcer til ajourføring og andre vedligeholdelsesopgaver. Det er derfor vigtigt, at disse stamdata skal have en kritisk efterspurgt masse, hvor andet data bedst egner sig til at blive stillet til rådighed på koordinering mellem fagdomæner. Til sidstnævnte kan Koncern IT med fordel tage en proaktiv rolle i governance omkring datadeling domænerne imellem.

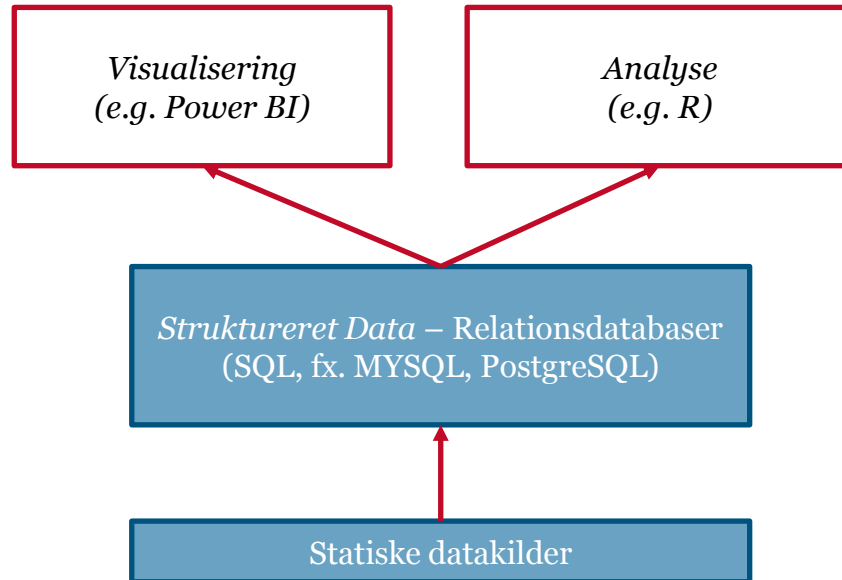
Appendiks

1. Note vedr. behandling af ubearbejdet data

Hvis analysearbejdet er bundet i ubearbejdet data stiger kompleksiteten i form af valg af platforme

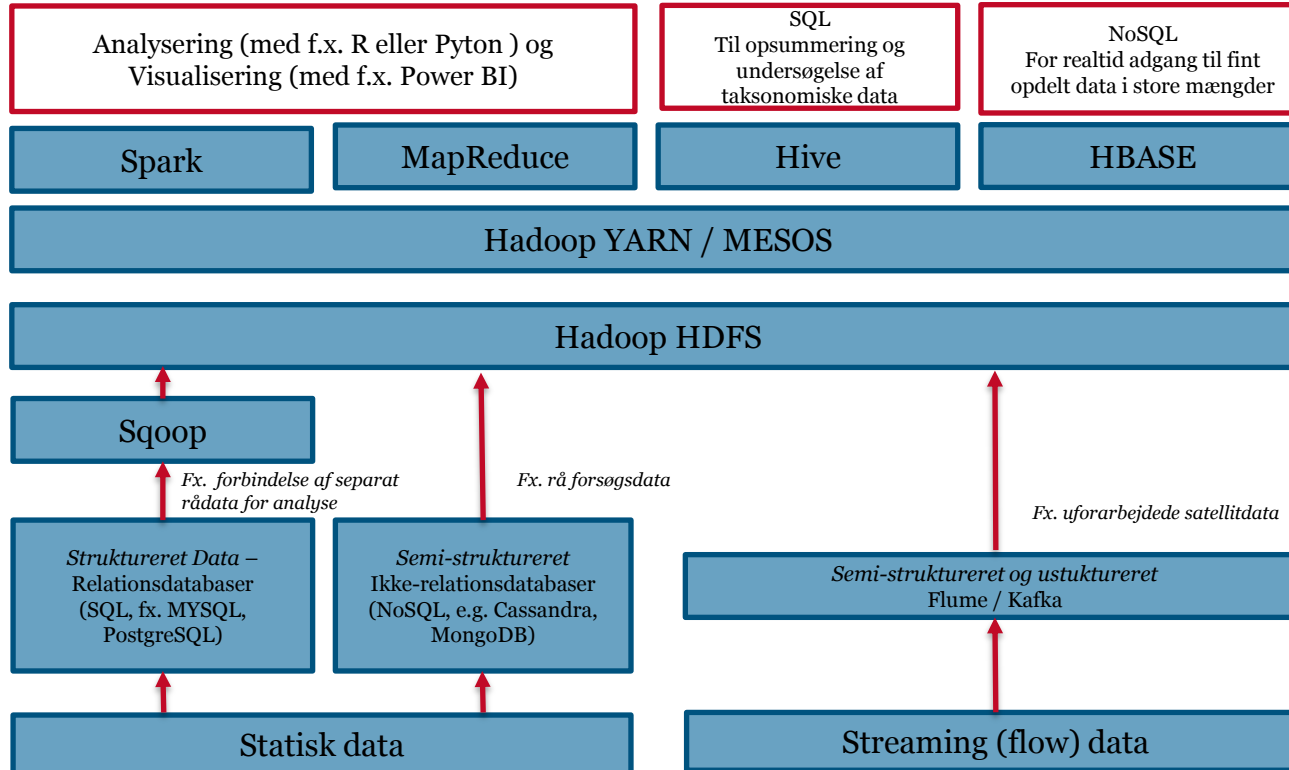
- Den påkrævede funktionalitet for mere avancerede opgaver er afhængig af typen af datakilder. For flere fagområder kan der være en business case på at modtage rådata (ex. IoT-data) og selv stå for strukturering. For planteområdet, eksempelvis, søger man bearbejdet GIS-data men kan ende med at skulle behandle mere rå og ubearbejdet data.
- Arbejdet med at behandle rådata er en særskilt disciplin, og åbner for en ny række tanker vedr. valg af platforme. Alt afhængigt om data hentes statisk eller automatisk bliver født ind i et datawarehouse, øges kompleksiteten og påkrævede platforme.

Dataarkitektur for struktureret (bearbejdet) data



Struktureret data kan placeres i en SQL database. Dvs. data med relationer og veldefinerede felter.

Mulig dataarkitektur for ustruktureret og semi-struktureret data



Semi-struktureret data er data der ikke kan placeres i et relationsdatabase, men har ikke desto videre noget struktur der gøre et nemt at bearbejde, det kan typisk findes i formater som CSV, XML og Json. NoSQL databaser bliver også betragtede som semistruktureret.

Ustruktureret data dækker alt andet data, det ville sige data der kan have intern struktur men som skal bearbejdes før det kan placeres i en database. Eksempler er billeder, uforarbejdede tekst (emails osv.)