



Your Professional Business Intelligence Partner

WHITE PAPER

10 grunner til å bruke Open Source BI

Terje Berg-Hansen, Mai 2013

Epost: berg-hansen@axenna.no

Web: www.axenna.no

Sammendrag

Fri programvare passer ideelt til BI-formål, og de største leverandørene (etter antall brukere) har typisk en open source-versjon av programvaren tilgjengelig for fri nedlasting. Fleksibilitet, kontroll, eierskap, standardisering og leverandør-uavhengighet er vel så viktig som kostnader når fri programvare vurderes. Fra første dag utføres testing og utvikling med fullstendige versjoner av programvaren, noe som gir kortere «time to market» for de ferdige løsningene. Deling av moduler og komponenter utviklet av andre BI-brukere supplerer leverandørene og fører til mindre utviklingskostnader og verdifull kompetanseoverføring. «Big data» er dominert av open source-løsninger som Hadoop, Cassandra, MongoDB osv., noe som gjør det enkelt å implementere semistrukturerte og ustrukturerte datamengder i BI-systemene.

Innledning

Bruken av Open Source BI øker i omfang, både i offentlig og privat virksomhet - og det er en rekke meget gode grunner til dette. I denne oppsummeringen skal jeg foreta en gjennomgang av 10 hovedmomenter som er ofte sitert når bedrifter begrunner hvorfor de valgte fri programvare til sine BI-løsninger. Som den våkne leser allerede har skjønnet, kommer jeg til å bruke begrepene Open Source Software og Fri Programvare som synonymmer i denne fremstillingen, og lar dette inkludere både gratis og kommersiell programvare, så lenge lisensieringen ivaretar de 4 frihetene som kjennetegner fri programvare (se punkt 6 nedenfor).

1 Fri programvare er ideelt til BI

Det er en generell sannhet at programvare som brukes til utvikling ikke er differensierende for en bedrift, siden alle kan bruke den samme programvaren. En bedrift skiller seg ut ved *måten* programvaren blir brukt på, og ikke gjennom hvilken programvare som brukes. Dette gjør Open Source-modellen ideell for programvare som brukes til å utvikle bl.a. BI-løsninger. Delte utviklingskostnader, gode distribusjonsmodeller, og stadig forbedret kvalitet på verktøyene som brukes gagnar alle.

En av hovedidéene bak fri programvare er *produksjon for å dekke behov* heller enn *produksjon for fortjeneste*. Programvare utvikles for å løse konkrete problemer, og siden de fleste problemer rammer flere aktører, tjener alle på at løsningene deles. Internett har gjort dette ikke bare mulig,

Document name: 10_grunner_Axenna_White_Paper	Status: Published	License: n/a	External reference: White Paper
Target: General Public	Date: 2013-05-31		Revised: n/a
Author: Terje Berg-Hansen, Axenna Norge	Approved: 2013-05-31		Page 2 of 7

men høyeffektivt. Utviklere samarbeider i effektive globale miljøer, noe som også fører til rask kompetanseoverføring mellom de aktive deltakerne i prosjektene.

2 Fleksibilitet

Fleksibilitet er det som oftest nevnes som grunn for bruk av fri programvare i BI-løsninger. Fri programvare gir økt fleksibilitet på flere områder:

a. Evaluering. Bedrifter og organisasjoner kan prøve ut verktøyene i sin egen takt, uten å være bundet av begrensede evalueringsperioder eller ufullstendige versjoner av programvaren.

b. Installering/utrulling. Etter en periode med utprøving og evaluering kan man enklere konvertere til en produksjonsfase, uten å måtte reinstallere en ny, fullisensiert programvarepakke, e.l. Man kan også installere programvaren fritt etter behov, på en eller flere servere, sentralisert eller desentralisert. Skalering innebærer ikke økte lisenskostnader. Siden lisenser, support og service ikke er bundet sammen, kan man ta avgjørelser om disse på et senere tidspunkt.

c. Tilpasningsmuligheter. Fri programvare kan brukes uten restriksjoner, på måter man kanskje ikke hadde forestilt seg ved prosjektstart. Programvaren kan også tilpasses ved skreddersøm hvis det er nødvendig for enkelte situasjoner, for eksempel ved bruk av nye databaser som krever nye typer tilkoblinger / drivere. Bugs man oppdager underveis kan fjernes uten å måtte gå den ofte lange veien om leverandøren. I verste tilfelle kan en bug være en showstopper, og den proprietære leverandøren ikke ville rette den opp før i neste versjon av programvaren. Med fri programvare har man full rett til å gå inn i kildekoden og rette opp feilen.

d. Mix & Match. Ved bruk av fri programvare har man full frihet og tilgang til å blande produkt-moduler fra ulike leverandører og prosjekter slik at man kan sette sammen sin egen BI-løsning etter behov. For eksempel brukes komponenten *Talend Open Studio (TOS)* til ETL-oppgaver i BI-suitene *SpagoBI* og *JasperServer*. Men det er en smal sak å bytte denne ut med *Pentaho DI (PDI, også kalt Kettle)* hvis man skulle foretrekke denne fremfor TOS.

3 Standardisering

Programvare som følger internasjonale standarder er viktig innenfor de fleste områder, men det er

Document name: 10_grunner_Axenna_White_Paper	Status: Published	License: n/a	External reference: White Paper
Target: General Public	Date: 2013-05-31		Revised: n/a
Author: Terje Berg-Hansen, Axenna Norge	Approved: 2013-05-31		Page 3 of 7

prekært innenfor BI, og spesielt innenfor data integrasjons-delen og datavarehus-delen av BI. Effektiv ETL – ekstrahering, transformering og lasting av data – krever at programvaren er kompatibel med eksisterende standarder for lagring av data.

De fleste proprietære programvareleverandører hevder at de følger etablerte standarder – men det er vanskelig å vite akkurat i hvilken grad, siden kildekoden er lukket, og siden leverandørene gjerne vil differensiere ved å tilby *litt mer* enn standard. Dette *litt mer* kan fort skape problemer når “legoklossene” skal kobles sammen, og noen klosser er litt større enn de andre.

Med fri programvare kan alle sjekke om noe som hevdes å følge en standard, faktisk gjør det. Og slike sjekker foretas kontinuerlig, av utviklere og sluttbrukere over hele verden. Med proprietær programvare er dette forretningshemmeligheter, og man må stole på leverandøren inntil eventuelt det motsatte er bevist. Muligheten til å saksøke en leverandør bidrar dessverre ikke til å fullføre prosjektet innen den gitte tidsrammen, tvert imot kan innledende juridiske manøvre lett forlenge prosjektperioden. Studier har også vist at innsynet som frie programvare-prosjekter gir, ansporer utviklere til å skrive kode som er mer robust og lettere å vedlikeholde.

4 Leverandør-uavhengighet

Leverandøravhengighet (såkalt vendor lock-in) skaper problemer på flere områder. I en proprietær leverandørverden er det gunstig for leverandører å søke å skape så stor avhengighet som mulig til egne produkter og tjenester. Intern avhengighet mellom en leverandørs produkter skaper gjerne i tillegg det som kalles teknologi lock-in, hvor kjøp av ett produkt innebærer at man i større eller mindre grad er låst til å bruke leverandørens produkter på beslektede områder.

Lisensieringen som brukes av fri programvareløsninger er nøkkelen til uavhengighet fra leverandører og teknologier. Med frie lisenser som GPL/LGPL o.a. er det ingen bindinger til bruk av en spesiell leverandør for kjøp av tjenester, supportavtaler eller liknende knyttet til produktet.

En annen trygghet ved fri programvare-modellen ligger i at prosjekter med et aktivt nettsamfunn av utviklere har stor sannsynlighet for å overleve dersom selskapet som har utviklet programvaren ikke skulle gjøre det. Leverandøren tar ikke med seg sine hemmeligheter i graven, noe som i verste fall kan skje med proprietære leverandører.

Document name: 10_grunner_Axenna_White_Paper	Status: Published	License: n/a	External reference: White Paper
Target: General Public	Date: 2013-05-31		Revised: n/a
Author: Terje Berg-Hansen, Axenna Norge	Approved: 2013-05-31		Page 4 of 7

5 Optimal pris

Oppstartskostnadene for proprietære BI-prosjekter kan være eksepsjonelt høye, mens de kan være tilnærmet lik null ved bruk av Open Source BI. Dette gjør BI med fri programvare til et attraktivt alternativ for små- og mellomstore bedrifter, organisasjoner og offentlige etater. Men kostnadsaspektet omfatter hele livssyklusen, inkludert skalering, support, arbeid med tilpasninger og bug-fjerning osv. Også her kommer Open Source BI svært godt ut. De fleste kommersielle leverandører av Open Source BI, som Pentaho, Jaspersoft, Talend og Engineering Group (SpagoBI) priser supportavtaler, tjenester og enterprise versjoner av programvaren etter modeller som gjør det svært mye rimeligere å skalere løsningen, enn hva som er tilfelle med mange proprietære leverandører, hvor lisenser ofte gjelder per antall CPU/Servere e.l.

IT-avdelinger har mindre budsjettmessig rom for prøving og feiling enn tidligere, og de høye proprietære BI-lisenskostnadene kombinert med det store antallet komponenter som inngår i en effektiv BI-løsning (Forrester telte mer enn 40 i 2009) gjør tradisjonell, proprietær BI lite aktuelt for mange bedrifter og organisasjoner, og Open Source BI tilsvarende mer attraktivt.

6 Kontroll over systemer og kode

Det er ofte hevdet av friprog-tilhengere at det ikke er kostnadsbesparelsene som er det viktigste aspektet ved fri programvare, men at brukeren har full kontroll over systemene og kildekode. De fire frihetene som ligger innbakt i begrepet Fri programvare innebærer:

1. Rett til å bruke programvaren, til valgfritt formål
2. Rett til å distribuere kopier av programvaren
3. Rett til å undersøke, endre og tilpasse programvaren
4. Rett til å distribuere endringene av programvaren

Dette gir en helt annen kontroll over programvaren, enn det som ligger innbakt i lisenser for proprietær BI-programvare.

Document name: 10_grunner_Axenna_White_Paper	Status: Published	License: n/a	External reference: White Paper
Target: General Public	Date: 2013-05-31		Revised: n/a
Author: Terje Berg-Hansen, Axenna Norge	Approved: 2013-05-31		Page 5 of 7

7 Eierskap til systemer og kode

Lisensiert proprietær programvare er leid, ikke eid. Det begrenser både hva du kan bruke programvaren til, og hvor lenge du kan ta vare på den. Fri programvare er eid, og kombinert med de fire frihetene fra forrige punkt gir dette økt fleksibilitet, trygghet og kontinuitet. Dette gjelder ikke minst når innleide konsulenter skal videreutvikle programvaren, utvikle nye moduler osv. Proprietære leverandører har tradisjon for å kreve eierskap til koden de utvikler for sine kunder – med fri programvare beholder bedriften eierskapet til det som utvikles. I tillegg kan det deles via internett for testing, feilretting eller tilføring av tilleggsfunksjonalitet av andre bedrifter og utviklere med tilsvarende behov. I mange tilfeller er modulen allerede utviklet av andre, og det er nok for dine konsulenter å videreutvikle eller endre deler av det som er laget.

8 Deling av ressurser, løsninger og kompetanse

De største aktørene på BI-markedet (etter antall brukere) tilbyr en open source-versjon av programvaren for fri nedlasting – så som Pentaho, Jaspersoft, Talend og SpagoBI. De fleste av disse tilbyr en betalt «enterprise-versjon» med tilleggsfunksjonalitet eller bundlede supportavtaler, mens f.eks. SpagoBI leverer en komplett løsning som er 100% fri og gratis. Det er utstrakt gjenbruk av komponenter mellom disse aktørene, og de fleste BI-løsningene er Java-baserte. Dette gjør at f.eks. både SpagoBI og Jaspersofts løsninger leveres med Talends Open Studio til ETL-oppgaver. Det gjør også at Jaspersofts rapporteringsmodul Ireport / Jasperstudio er integrert i både SpagoBI og Pentahos løsninger.

I tillegg finnes det markedsplasser for salg og utveksling av egenproduserte komponenter. Jasperforge og Talendforge er to slike, hvor aktører fra hele verden utveksler programvare og kompetanse.

9 Raskere utviklingstakt, større utviklermiljøer

Fri programvare utvikles gjerne med nettbaserte prosjektmodeller hvor ny funksjonalitet, testing, bug fixing, support osv. foregår med så mange involverte personer fra så mange deler av kloden at det er umulig for ett selskap, uansett hvor stort det måtte være, å holde tritt med de største Open Source prosjektene. Dette har en selvforsterkende effekt – jo raskere utviklingen går, jo flere vil

Document name: 10_grunner_Axenna_White_Paper	Status: Published	License: n/a	External reference: White Paper
Target: General Public	Date: 2013-05-31		Revised: n/a
Author: Terje Berg-Hansen, Axenna Norge	Approved: 2013-05-31		Page 6 of 7

bruke programvaren, og jo flere som bruker den, jo flere vil være med å utvikle den. Også der hvor kjerneutviklerne tilhører ett friprog-selskap, er det som regel i tillegg et nettsamfunn av andre utviklere, modulbyggere, testere osv som akselererer utviklingen, bedrer kvaliteten osv.

10 Elefanten Hadoop og vennene hans

Sist, men absolutt ikke minst: Apache Hadoop er blitt den rådende løsningen innen fenomenet «Big Data». Og Hadoop er Open Source. Det vil si at en rekke andre Open Source løsninger er godt integrert med Hadoop:

- Databaser (MongoDB, Cassandra)
- Datavarehus (Hive)
- Data mining (Pig, Weka)
- Dataintegrerings-løsninger (Pentaho, Talend)
- BI-servere (Pentaho, Jaspersoft, SpagoBI)

Hadoop implementerer den kjernefunksjonaliteten som er i sentrum i de fleste moderne Enterprise Datavarehus: nettsky-arkitektur, MPP, direkte database-analyser, håndtering av «mixed workload» og et hybrid lagrings-system. Hadoop representerer på mange måter en redefinerings av det tradisjonelle datavarehuset for tilpasning til nettsky-sentriske forretningsmodeller som krever raske svar på avanserte analyser av store datamengder.

Når det gjelder «Big Data» er Open Source så sentralt at det blir perifert å bruke proprietære løsninger. Og «Big Data»-måten å tenke lagring, integrering, analysering og skalering på øker i utbredelse på en måte som gjør det sannsynlig at begrepet *paradigme-skifte* vil dukke opp oftere i debatten fremover. Og i dette nye paradigmet vil Open Source BI ha en naturlig mer sentral plass enn i det gamle, proprietært dominerte regimet.

Document name: 10_grunner_Axenna_White_Paper	Status: Published	License: n/a	External reference: White Paper
Target: General Public	Date: 2013-05-31		Revised: n/a
Author: Terje Berg-Hansen, Axenna Norge	Approved: 2013-05-31		Page 7 of 7