

Kjetil Sander presenterer:



Big Data



Hva er Big Data og hvordan kan selskapet ta i bruk Big Data for å øke sin lønnsomhet, skape innovasjon og oppnå nye konkurransefortrinn?

Skrevet av:
Kjetil Sander

Utgitt av:
eStudie.no

Revisjon:
1.0 (Sept. 2017)

Innhold

1 Big Data	4
1.1 Hva er Big Data?	4
1.2 Grunnlag for stordata	6
1.3 Hva er driverne bak Big Data?	7
1.4 Kilder til Big Data	8
1.5 Direkte datainnsamling	8
1.6 Automatisk datainnsamling	9
1.7 Frivillig dataproduksjon	9
1.8 Åpne data	10
1.9 Egenskaper ved Big Data	10
1.9.1 Datastørrelse	10
1.9.2 De tre v-ene	10
1.9.3 De fem v-ene	11
1.9.4 Flere egenskaper ved stordata	11
1.10 Hvordan utnytte Big Data?	12
1.10.1 En kilde til innsikt og økt lønnsomhet	12
1.10.2 Bare fantasien setter grensen	13
1.10.3 Faktabaserte beslutninger	13
1.10.4 Hvordan gjøre rådata til "gull"?	14
1.11 Data mining gir nye muligheter	15
1.11.1 Data mining	16
1.11.2 Datautvinning	18
1.12 Hvordan brukes Big Data idag?	19
1.12.1 Big Data i bruk blant internettbaserte selskaper	19
1.12.2 Big Data innen forsikring og kredittvurdering	20
1.12.3 Big Data på helseområdet	21
1.12.4 Big Data innen politi, sikkerhet og etterretning	22
1.13 Verdikjeden til Big Data	23
1.13.1 Innsamling av data	23
1.13.2 Lagring og aggregering	27
1.13.3 Analyse	28
1.13.4 Brukere og bruksområder	30
1.14 Personvernprinsipper under press	32
1.14.1 Enorme mengder data sammenstilles	33
1.14.2 Ulemper for enkeltmenneskene	33
1.14.3 Ti sentrale personvern-utfordringer:	34
1.14.4 Mulig å respektere personvernet	36
1.15 Innebygd personvern	37
1.15.1 Vær i forkant, forebygg fremfor å reparere	37
1.15.2 Gjør personvern til standardinnstilling	38
1.15.3 Bygg personvern inn i designet	38
1.15.4 Skap full funksjonalitet	39
1.15.5 Ivareta informasjonssikkerheten fra start til slutt	40
1.15.6 Vis åpenhet	41
1.15.7 Respekter brukerens personvern	41
1.16 Big Data og sikkerhet	42
1.16.1 1. Sikker infrastruktur	42

1.16.2	2. Beskytt rammeverket for distribuert programmering	43
1.16.3	3. Sikring av Non-relasjonsdata	44
1.16.4	4. Sikker lagring og transaksjonslogger	44
1.16.5	5. Endpoint Filtrering og validering	45
1.16.6	6. Real-Time Compliance and Security Monitoring	45
1.16.7	7. Big Data Kryptografi	45
1.16.8	8. Adgangskontroll	46
1.17	Big Data plan.....	46
1.17.1	Lag en big data plan.....	46
1.17.2	Situasjonsanalyse.....	47
1.17.3	Velg noen få satsningområder	48
1.17.4	Rammebetingelser	49
1.17.5	Mål	49
1.17.6	Strategi.....	49
1.17.7	Program	50
1.17.8	Implementering	50
1.17.9	Tidsplan og budsjett	50
1.17.10	Kontroll og evaluering	50
1.17.11	Datakilder	51
1.17.12	Analytiske modeller	52
1.17.13	Verktøy	52

1 Big Data

1.1 Hva er Big Data?

La oss starte med EU-kommisjonens rådgivende organ i personvernspørsmål sin definisjon av Big Data:

"Data refererer til den enorme økningen i tilgang til, og automatiserte bruk av, opplysninger: det refererer til gigantiske mengder digitale data som er kontrollert av selskap, myndigheter og andre store organisasjoner, og som gjøres til gjenstand for omfattende analyse ved bruk av algoritmer. Big Data kan bli brukt til å identifisere generelle trender og sammenhenger, men kan også bli benyttet slik at det berører enkeltindivider direkte"

Dette er egentlig ingen definisjon, men en kort beskrivelse av hva Big Data er.

Det første forsøket på en definisjon av Big Data ble lansert i 2001:

"Big Data er data som kjennetegnes ved tre V'er, Volume (volum), Variety (variasjon) og Velocity (hastighet)".

- **Volume (volum):** Tilgjengelig data øker i omfang – IBM estimerer at det innen 2020 kommer til å genereres 2,3 trillioner gigabyte data hver dag, som tilsvarer 107 millioner blue-ray disketter eller 4,2 milliarder fulle CD'er
- **Variety (variasjon):** Ulike typer data er tilgjengelig fra en rekke ulike datakilder - 400 millioner Twitter-meldinger sendes hver dag, og stadig flere datakilder kobles opp mot internett og genererer informasjon
- **Velocity (hastighet):** Data er tilgjengelig raskere enn før og vi ønsker at dataene er tilgjengelige i tilnærmet sanntid

IBM har valgt å ta med enda en V i sin definisjon av Big Data:

- **Veracity (pålitelighet):** usikkerhet knyttet til kvalitet i dataene, og viktigheten av å kunne stole på dataene.