

Sikker datasjø

En introduksjon til forvaltningsrammeverket



STAVANGER KOMMUNE

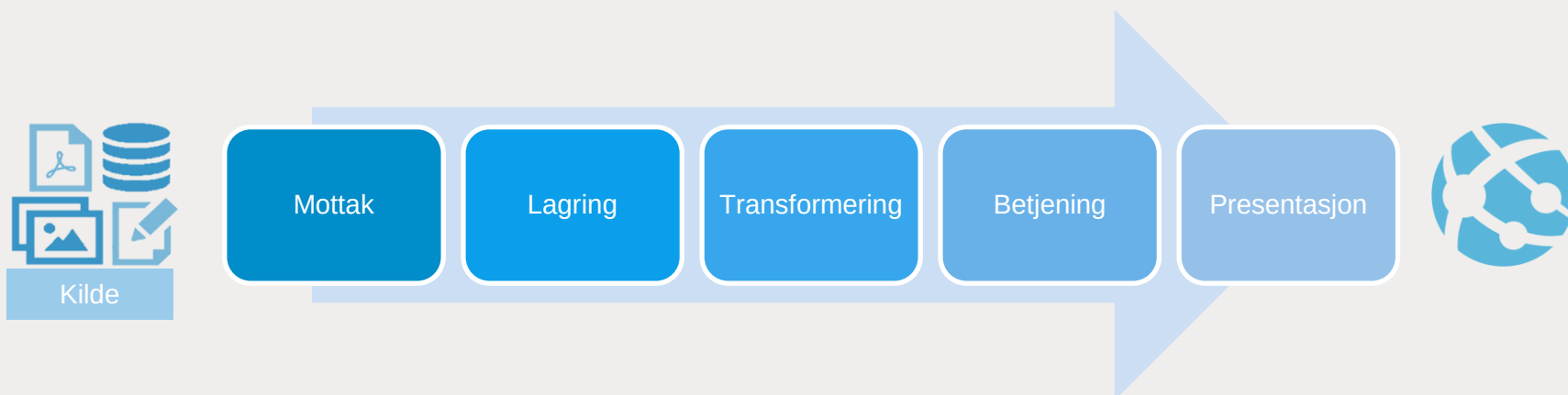
Introduksjon

- Stavanger kommunes strategiske mål er å få kontroll på og eierskap til egne data. Dette gjelder alle typer data fra kommunens ulike fagsystemer, samt andre typer data som for eksempel sensor data. Både sensitive og ikke-sensitive data. Stavanger kommune har etablert en sikker datasjø for å få alle data samlet i ett felles område. I tillegg til å gi kontroll og eierskap over dataene, vil datasjøen også gi verdi ved å legge til rette for analyse på og innsikt i de store og varierte datamengdene kommunen har.
- For å tilrettelegge for denne type analyse og innsikt, har Stavanger kommune valgt Microsoft Azure dataplattformen med tilhørende komponenter. Løsningen vil åpne for håndtering av alle typer data, analyse på tvers av fagsystemer, bruk av data til andre og nye formål som for eksempel innsyn, smartbyutvikling og bedre tilgjengeliggjøring av åpne data. Samtidig som denne datasjøen vil bistå kommunen med å overholde sine plikter i forhold til personvernlovgivningen.

Formål

- Forvaltnings- og sikkerhetsrammeverket gir en oversikt over Stavanger kommunes regler og retningslinjer ved bruk av komponenter og tjenester på Microsoft Azure plattformen.
- Rammeverket dekker hvordan man må gå fram for å analysere kilder, dokumentere dem og laste data opp i datasjøen, samt retningslinjer for hvordan data kan aggregeres, anonymiseres og ellers tilrettelegges for analyse og innsikt for sluttbrukere og mottagende systemer.

Oversikt



- Flyt av data i en datasjø går som regel igjennom 6 steg eller stadier fra kildesystemer fram til konsum hos mottager. Samtidig er det ofte gunstig å se på dataflyten som bevegelser i separate lag, og dette rammeverket knytter tjenestene som skal brukes i datasjøen til lag. Sikkerheten i datasjøen er koblet til de ulike tjenestene i de spesifikke lagene de skal benyttes. Enkelte tjenester vil ha en rolle i flere av lagene.

De tre delene av datasjøen

■ Del 1 – Hente data fra kilde og lagre i datasjø

- Kildesystemer – Analyse av kildesystem for å hvordan best mulig håndtere kopiering av data i datasjø, dokumentasjon og videre bruk av data til ulike formål.
- Mottak - Tjenester som fungerer som inntak til datasjøen
- Lagring - Hvordan lagre, og sikre rådata i datasjø

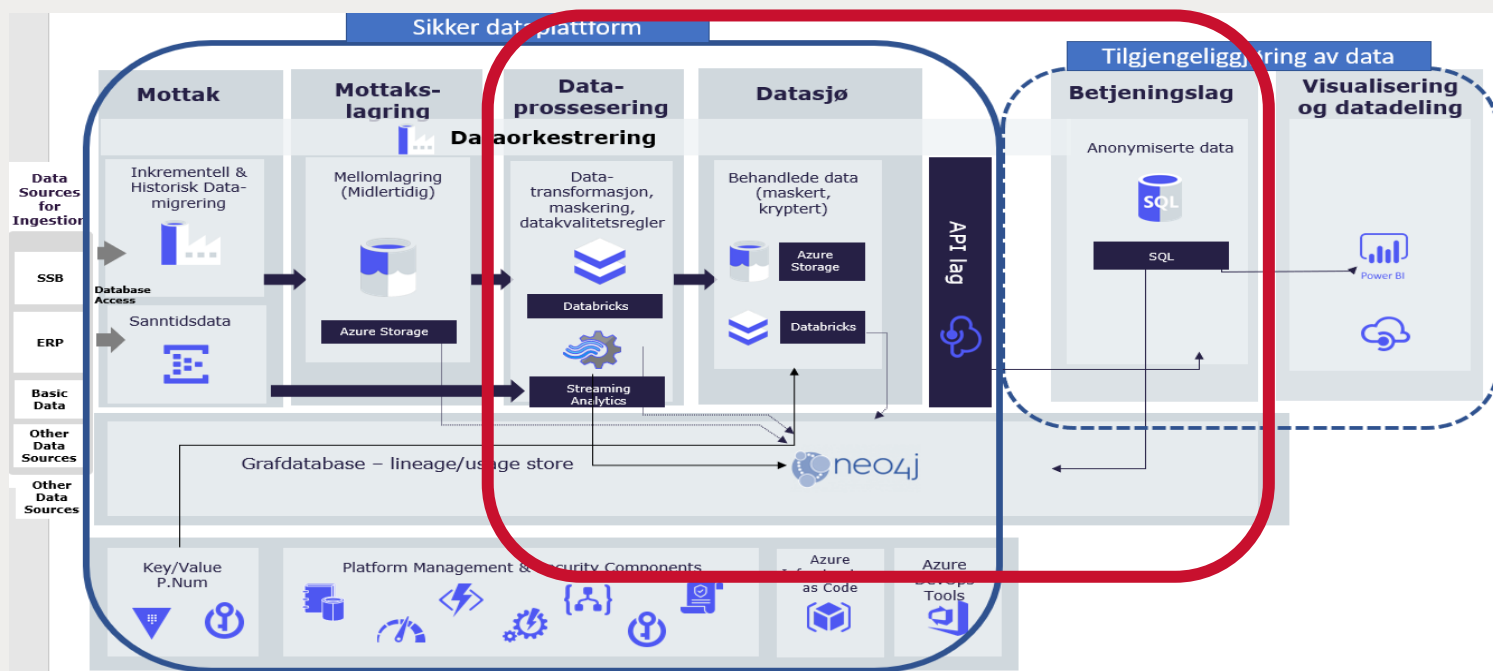


- Data hentes fra kildesystem uten å endres og lagres dermed som rådata i datasjø.
- Ved å ha gode rutiner for analyse av og dokumentasjon av kildesystemer vil både rett sikkerhetsnivå kunne legges på de ulike kildene, og man sikrer at persondata kan håndteres på riktig måte. Alt mottak av persondata loggføres i en egen database. Forvaltning av tilgang til rådata og prinsipper og rutiner for loggføring av bruk av dataene sikrer sporbarhet og kontroll.

De tre delene av datasjøen

■ Del 2 – Bruke data lagret i datasjø

- Transformering – Hvordan lese og transformere data fra datasjø
- Betjening – Hvordan lagre data på best mulig måte for de ulike behovene/formålene
- Presentasjon – Presentasjon/Visualisering av data til mottakere. Også kalt applikasjon i enkelte skisser

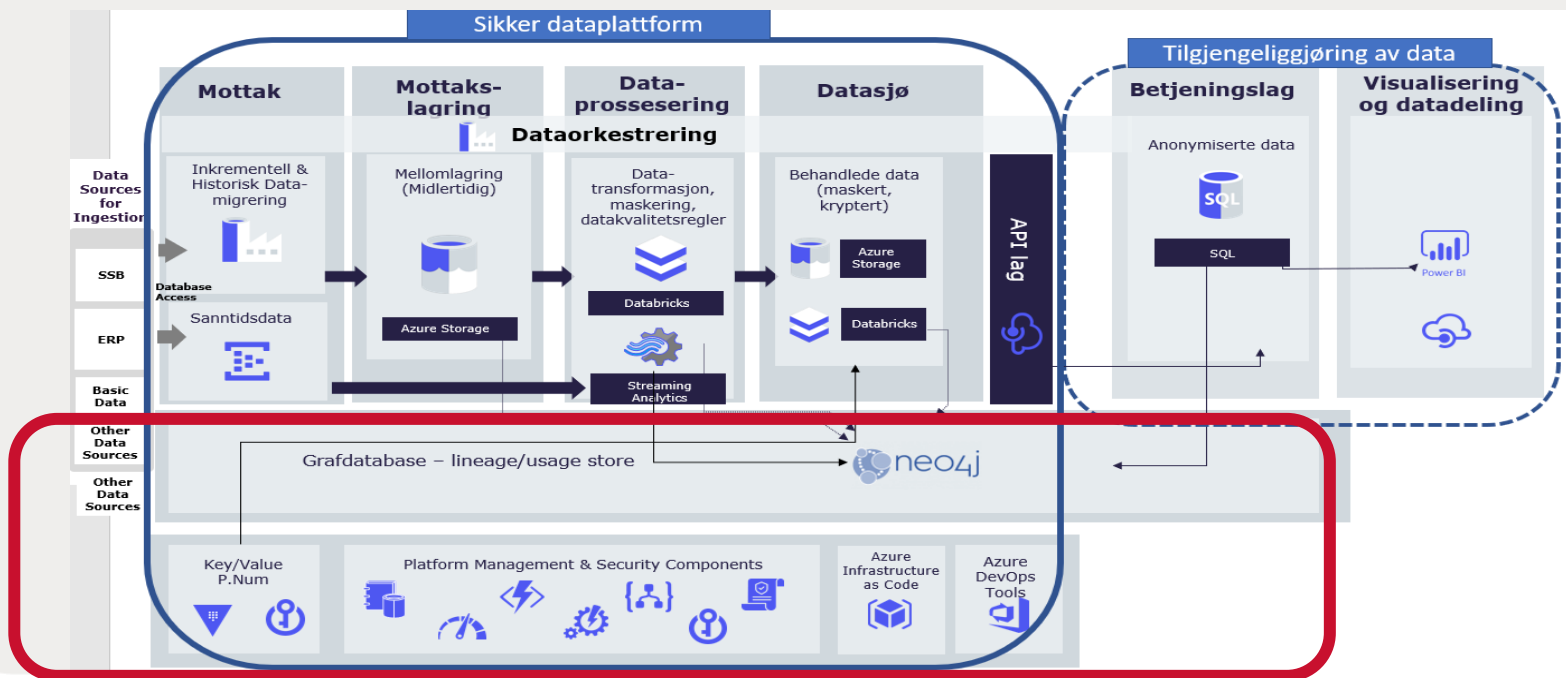


- Transformering av data er en operasjon som både kan innebære anonymisering av sensitive data, aggregering for å levere data til bruk for generell analyse eller sammenstille data fra ulike systemer slik at det kan gjøres analyse på tvers uten behov for dataforskningskompetanse. All behandling av personsensitive data vil loggføres i en egen database.
- Betjening er å gjøre dataene tilgjengelig for konsum av ulike mottagere – det være seg andre datasystemer eller sluttbrukere. Tilgangskontroll, kryptering, behov for anonymisering og andre sikkerhetshensyn spiller inn på hvilke tjenester som vil bli valgt til betjening avhengig av typen data.

De tre delene av datasjøen

■ Del 3 – Felles Funksjoner

- Datakatalog – Dokumentasjon av alle data i de ulike lagringsmediene
- Brukerloggdatabase – Register over hvilke personers data som er i datasjøen, hvor de er lagret og hva de har blitt brukt til.
- Overvåking – Hvordan overvåke de ulike tjenestene i plattformen.

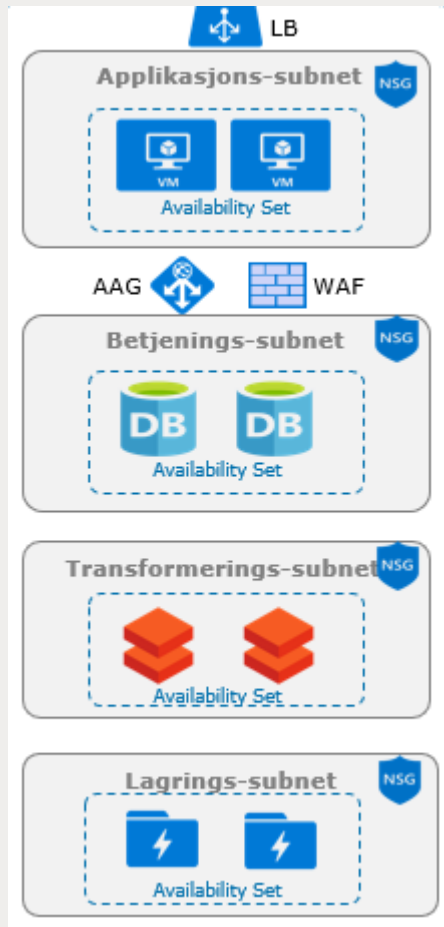


- Datakatalogen er nøkkelen til datasjøen, her vil brukere kunne søke etter relevante data for å drive selvbetjent analyse. Men her vil kommunen også sørge for den kontrollen over dataflyt og sporbarhet som personvernlovgivningen krever.
- Brukerloggdatabasen sikrer at kommunen etterlever sine plikter til personvernlovgivningen relatert til datasubjektets rettigheter.
- Uansett hvor god sikkerhet som er bygget inn i datasjøens arkitektur, så vil ikke det bety noe hvis man ikke har rutiner og verktøy til å overvåke og følge opp systemene når de er i bruk.

Sikkerhet

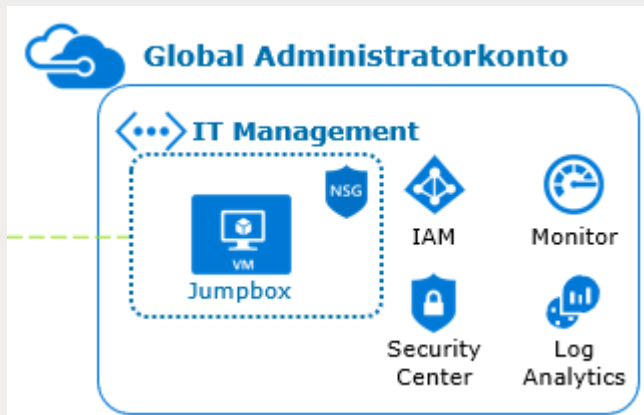
- Sikring av data i en skytjeneste er like viktig som sikring lokalt. Microsoft har etablert en hel rekke verktøy i Azure for å øke sikkerheten, og de fleste av disse er tatt i bruk på Stavanger kommunes datasjø. I tillegg har man innført en egen tjeneste for kryptering av sensitive data. Alt er satt opp etter beste praksis. De 6 lagene i datasjøen kan gjenkjennes i skissen under, om enn i et annet oppsett enn i selve løsningsarkitekturen som ble vist på foregående lysbilder.

Sikkerhet



- Ved å legge hvert lag i datasjøen i eget sub-nett med adgangskontroll får man meget god kontroll på data i alle stadier.
- WAF – web application firewall sikrer at ikke potensielle angripere kan utnytte de vanligste angrepsmetodene
- Separate miljø skal sikre at innleide konsulenter ikke har tilgang til data de ikke skal se

Sikkerhet



- Dedikerte sikkerhetsansvarlige har tilgang til en rekke verktøy for både å sette opp alarmer ved mistenkte sikkerhetsbrudd, men også avanserte analyseverktøy for å etterforske potensielle brudd.
- Prinsippet om delegering av myndighet skal gjelde. De som sitter som sikkerhetsansvarlige har ikke tilgang til data, og motsatt.