

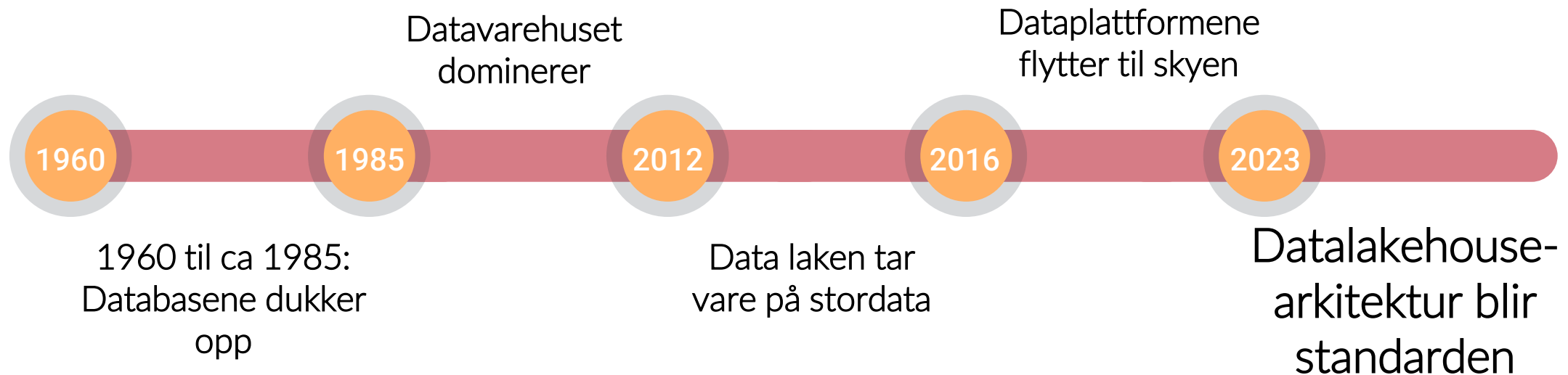


7 spådommer for data og analyse i 2023

Les alle spådommene på
glitni.no/fagartikler/

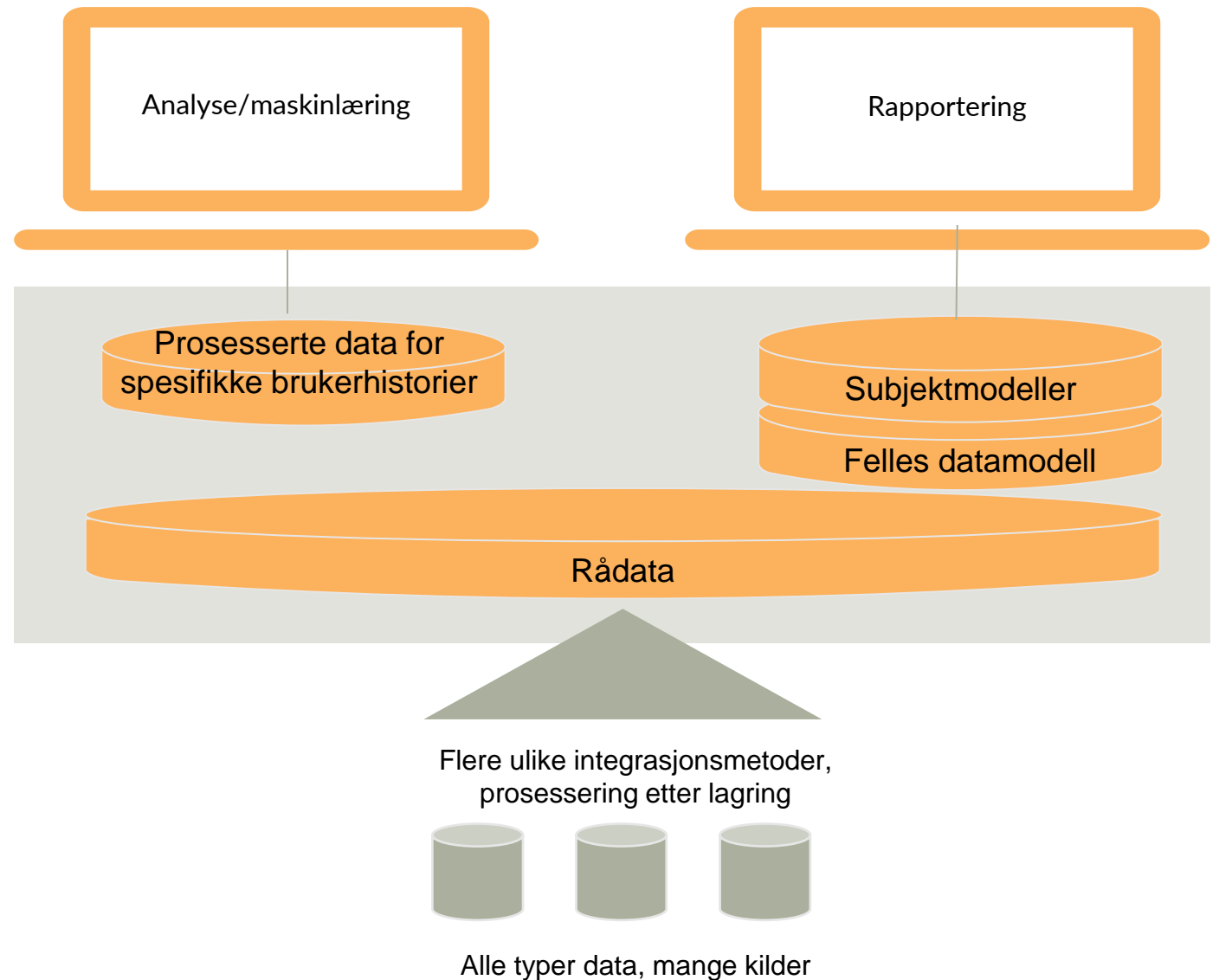
1. Arkitektur: Datalakehousene overtar!

2023 blir året da datalakehouse-arkitektur blir standarden for alle større virksomheter



Mindre dataflytting

Data lakehouse-modellen, der vi har en data lake med rådata og logiske datavarehus-lag på toppen, bidrar til å **minimere dataflytting og -lagring.**



Vi kan endre og slette data

Delta Lake-filformatet (og alternativene Hudi og Iceberg) gjør at vi også får **logiske databaseoperasjoner** til Data lake-filer.

Litt teknisk har de støtte for ACID, som gjør at vi blant annet kan både endre og slette data - og det må vi kunne for å f.eks imøtekomme GDPR-krav.



Prosesseringsmotorene gir kraft

Teknologi som Databricks og Snowflake, kjørt på skyinfrastruktur, gjør det mye enklere å få dette til å spinne rundt.

Endelig kan vi bruke energien vår på å **skape verdi** fra dataene, fremfor å jobbe med å provisjonere hardware, konfigurere for ytelse eller utføre oppgraderinger.



Har dere ikke noe fra før?

Om dere har tydelige use case dere ønsker å realisere som omfatter **både rapportering og data science** er nok datalakehouse-arkitektur verdt en diskusjon

Mye teknisk gjeld og lite dokumentasjon?

Det tar tid å flytte det gamle. Vi anbefaler å vurdere hvilke deler av løsningen som bør bygges på nytt, og å flytte inn i det nye gradvis.

2. Organisering:

Alle vil ha dataplattformteam!

Hvem skal videreutvikle fellestjenesten «dataplattform»?

Stadig flere selskap skjønner at utvikling av dataplattform ikke er det samme som å utvikle data-pipelines, rapporter og andre tjenester på en skyplattform eller i egen infrastruktur.

Skyen fikser mange av infrastrukturoppgavene, men vi bruker data på mer avanserte måter

Plattformteamet må derfor ha god forståelse for:

- △ forretningskrav
- △ DevOps/DataOps/MLOps
- △ infrastruktur
- △ integrasjonsmetoder og -grensesnitt
- △ dataflyt (ELT, Pub/Sub)
- △ modellering
- △ overvåkning
- △ datatesting (data observability)
- △ sikkerhet og tilgangsstyring

Super-personen som kan alt dette til
fingerspissene finnes naturlig nok ikke, så bygg
komplementær kompetanse i teamet.

Bruk gjerne også konsulentmarkedet for å få etablert teamet, plattformen og gode prosesser og rutiner.

Husk å bygge **intern kompetanse** på veien!

Kompetansen spres!

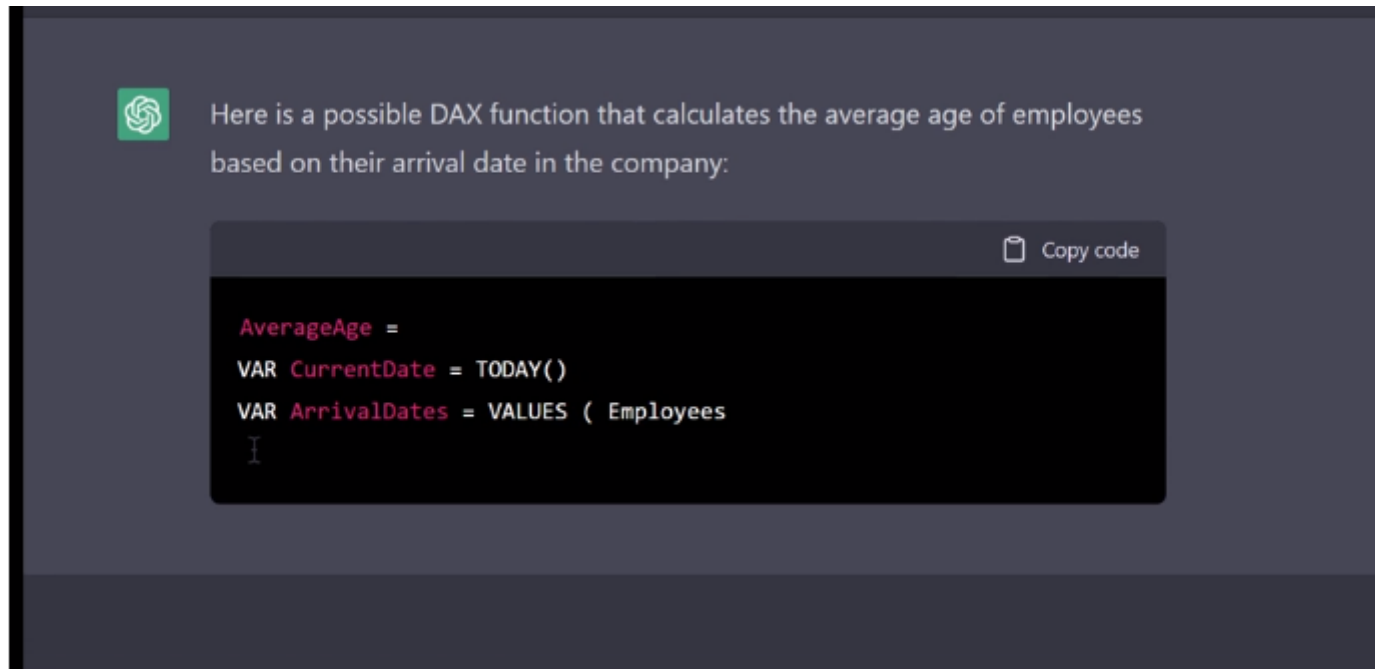
Mindre funding til start-ups og scale-ups, og kostnadsreduksjoner i større virksomheter gjør at mobiliteten til datafolket øker.

Konsekvensen av at noen av de mer modne selskapene kutter, er at kompetansen spres.

3. Dataflyt: datainntak på mange vis & data som kode

AI øker farten til utviklere (1/2)

AI vil effektivisere utvikling av transformasjonsjobber - og også benyttes innen dataprofilering og datavalidering.



The screenshot shows a dark-themed interface with a green OpenAI logo icon on the left. To its right, text reads: "Here is a possible DAX function that calculates the average age of employees based on their arrival date in the company:". Below this text is a dark code editor box containing the following DAX code:

```
AverageAge =  
VAR CurrentDate = TODAY()  
VAR ArrivalDates = VALUES ( Employees
```

 A cursor is visible at the end of the third line. In the top right corner of the code editor box, there is a "Copy code" button with a clipboard icon.

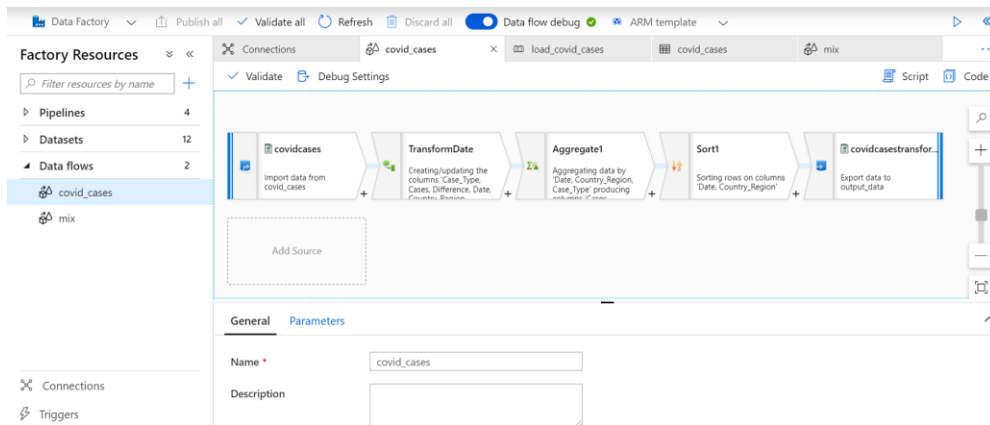
AI øker farten til utviklere (2/2)

Nå som vi kan i klartekst si f.eks at “gi meg en tabell som viser omsetning summert per måned, fordelt på produkt” og så få et forslag til SQL-kode, vil det for mange **gå betydelig fortere** å utvikle transformasjoner.

Dette vil bli bygget inn i de fleste verktøy der kode skrives. Power BI har allerede lansert støtte.

GUI vs kode for transformasjon: vi heier på kode! (1/2)

Verktøy som dbt får en stadig større utviklerskare. Andre sverger fortsatt til mer visuelle grensesnitt med drag&drop.



```
-- Refunds have a negative amount, so the total amount should always be >= 0.  
-- Therefore return records where this isn't true to make the test fail  
select  
  order_id,  
  sum(amount) as total_amount  
from {{ ref('fct_payments' )}}  
group by 1  
having not(total_amount >= 0)
```

GUI vs kode for transformasjon: vi heier på kode! (1/2)

Hva skal dere velge?

1. **Mindre organisasjoner** kan gjerne bruke verktøy som Data Factory og Alteryx - som representerer low/no code.
2. **Større organisasjoner** som har større utviklermiljøer bør i størst mulig grad benytte kode. I hvert fall for de offisielle dataflytene. Superbrukere i linja vil i stor grad også kunne bruke kodebaserte verktøy - og elske det.

Universelle semantiske lag - blir de tatt i bruk? (1/2)

Universelle semantiske lag ble lansert av dbt og Looker i 2022, og det blir spennende å se om konseptet brer om seg.

Metric 1

SOURCE Jaffle Shop Browse

370 rows | 10 seconds | 33.08 KB

Preview Display

View Format Filter

	DATE_MO...	A LOCATION...	123 REVENUE	123 REVENUE_...
0	2022-10-01T00:...	San Francisco	\$4,068	\$54,754
1	2022-10-01T00:...	Philadelphia	\$3,503	\$45,124
2	2022-10-01T00:...	New Orleans	\$3,532	\$41,415
3	2022-10-01T00:...	Chicago	\$5,804	\$61,881
4	2022-10-01T00:...	Brooklyn	\$5,220	\$72,033
5	2022-09-01T00:...	San Francisco	\$81,977	\$78,648
6	2022-09-01T00:...	Philadelphia	\$68,494	\$64,374
7	2022-09-01T00:...	New Orleans	\$63,217	\$59,060
8	2022-09-01T00:...	Chicago	\$92,668	\$87,840
9	2022-09-01T00:...	Brooklyn	\$108,481	\$102,726
10	2022-08-01T00:...	San Francisco	\$78,216	\$79,355
11	2022-08-01T00:...	Philadelphia	\$63,374	\$64,516

Page 1 of 31

metric_result 3

Universelle semantiske lag - blir de tatt i bruk? (2/2)

Formålet er å kunne tilgjengeliggjøre måltall på tvers av organisasjonen og sikre at alle bruker **de samme definisjonene og det samme datagrunnlaget**.

Datasett og presentasjon skiller lag, er annen måte å se på dette. Det er vel et forsøk på å fravriste Power BI rollen som verktøyet som fikser alt?

Råd: utforsk, og hør med de som har fått erfaring før dere går all-in

Ja takk, vi vil ha flere måter å få tak i data på

I 2023 blir vi mer vant til å si “ja takk” til flere måter å få data inn til en dataplattform på (batch, strømming osv).

Det betyr:

1. Data engineers må enten lære seg flere verktøy og arkitekturmønstre, eller samarbeide med andre utvikler- og integrasjonsteam for å få dataene inn.
2. Vi må forstå hvor dataene kommer fra, hvordan de er behandlet og hvilke forutsetninger man kan legge til grunn når dataene skal tolkes.

4. DataOps, MLOps og Data Governance: Data observability + MLOps fremfor PoC

Hvordan skal vi få stabile, sikre og effektive dataløsninger?

Svaret er kanskje **DataOps** og **MLOps**.

Det finnes mange Ops-begreper. Fellesnevneren er at det handler om å få prinsipper knyttet til DevOps+Smidig fra programvareutvikling til å møte dataverdenen.

Data observability blir noe data engineers begynner å prate om, men få virkelig får til

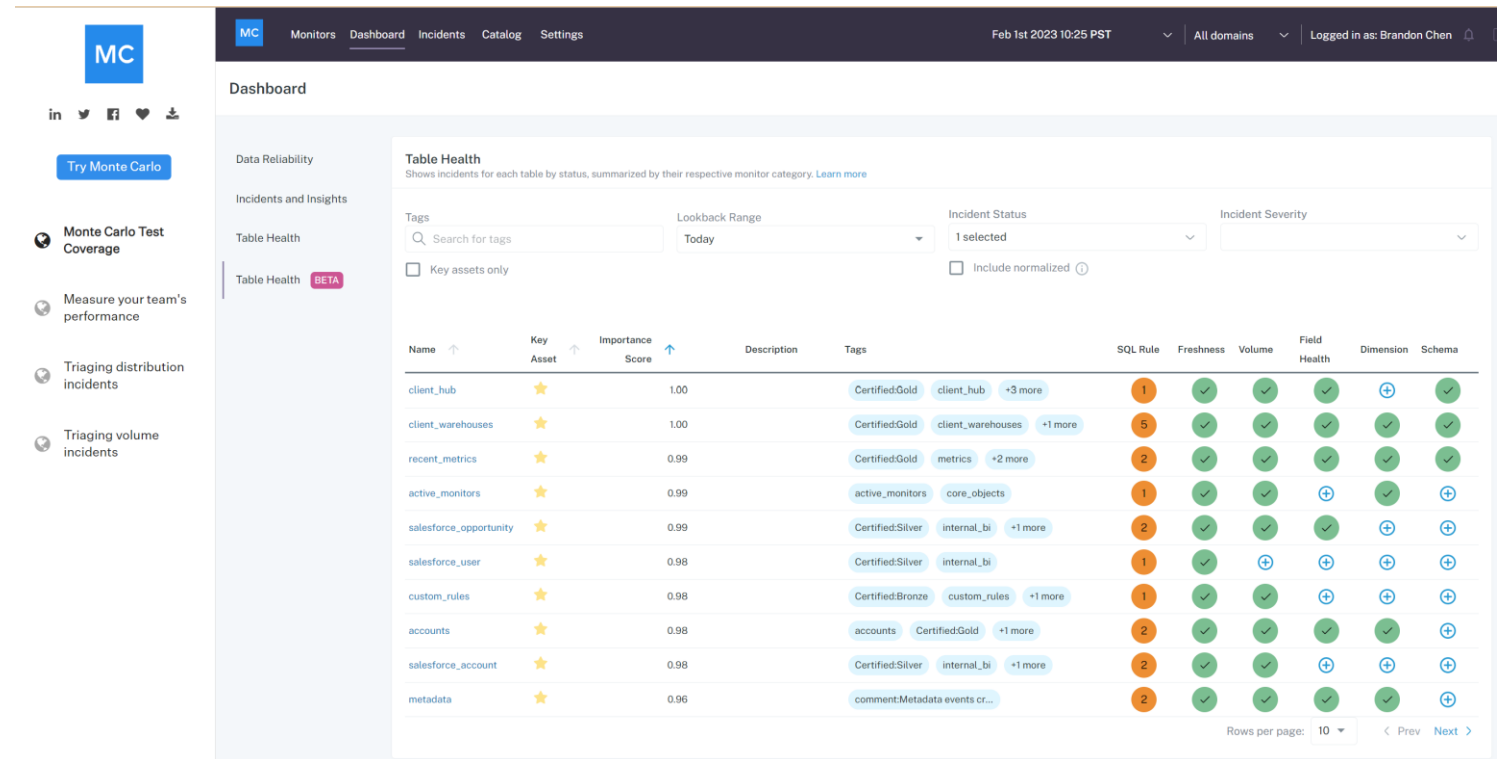
Data observability handler om å automatisere overvåkning av dataflyten, med vekt på

- **Data lineage:** hva som har skjedd med dataene frem til et punkt, mao hvordan flyten av transformasjoner har vært
- **Datakvalitet:** ulike logiske sjekker som viser om dataene har endret karakter utover det vi forventer

Data observability blir noe data engineers begynner å prate om, men få virkelig får til

Her modnes både verktøy og metoder over tid, men foreløpig er vi startgropen. I 2023 blir temaet satt mer på agendaen. Det er en start.

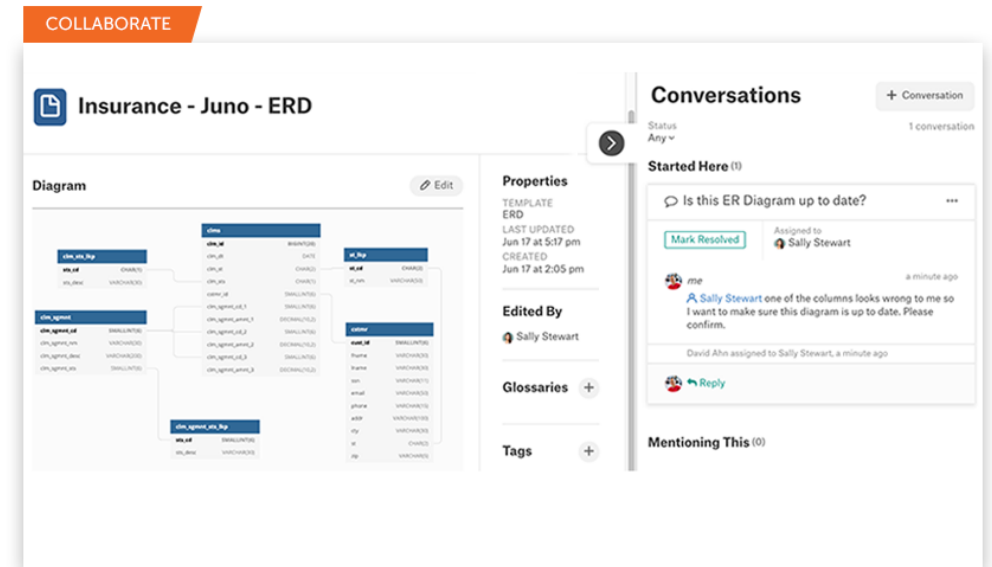
Bilde hentet fra montecarlo.com
19.2.2023



Name	Key Asset	Importance Score	Description	Tags	SQL Rule	Freshness	Volume	Field Health	Dimension	Schema
client_hub	★	1.00		Certified:Gold client_hub +3 more	1	✓	✓	✓	+	✓
client_warehouses	★	1.00		Certified:Gold client_warehouses +1 more	5	✓	✓	✓	✓	✓
recent_metrics	★	0.99		Certified:Gold metrics +2 more	2	✓	✓	✓	✓	✓
active_monitors	★	0.99		active_monitors core_objects	1	✓	✓	+	✓	+
salesforce_opportunity	★	0.99		Certified:Silver internal_bi +1 more	2	✓	✓	✓	+	+
salesforce_user	★	0.98		Certified:Silver internal_bi	1	✓	+	+	+	+
custom_rules	★	0.98		Certified:Bronze custom_rules +1 more	1	✓	✓	+	+	+
accounts	★	0.98		accounts Certified:Gold +1 more	2	✓	✓	✓	✓	+
salesforce_account	★	0.98		Certified:Silver internal_bi +1 more	2	✓	✓	+	+	+
metadata	★	0.96		comment:Metadata events cr...	2	✓	✓	✓	✓	+

Mye data, mange brukere, og mange ulike use case krever en datakatalog for å bevare kontrollen (1/2)

Jeg ønsker meg **automatisert generering og deling av metadata** gjennom hele dataflyten, tilgjengelig for alle som trenger de, med muligheter for sosial samhandling slik at vi kan lære av hverandre hvordan dataene bør brukes.



Bilde hentet fra alation.com 19.2.2023

Mye data, mange brukere, og mange ulike use case krever en datakatalog for å bevare kontrollen (2/2)

Datakataloger krever mye av mennesker og prosesser for å enes om definisjoner, eierskap og rutiner.

Den praktiske konsekvensen det nærmeste året er at datakatalog-initiativ **skaleres ned** til de mest sentrale data-domenene og holdes på et minimum. Og det er greit. Og kanskje er det også slik at vi kommer til å ha datakatalog-kapabiliteter **spredd på flere verktøy**.

Vi får mer oppmerksomhet rundt MLOps – fremfor ML-PoCs (1/3)

Vi ser fortsatt at implementering av løsninger som inneholder komponenter basert på AI har en del mangler. Dessverre kommer de fleste løsningene ikke forbi PoC-stadiet.

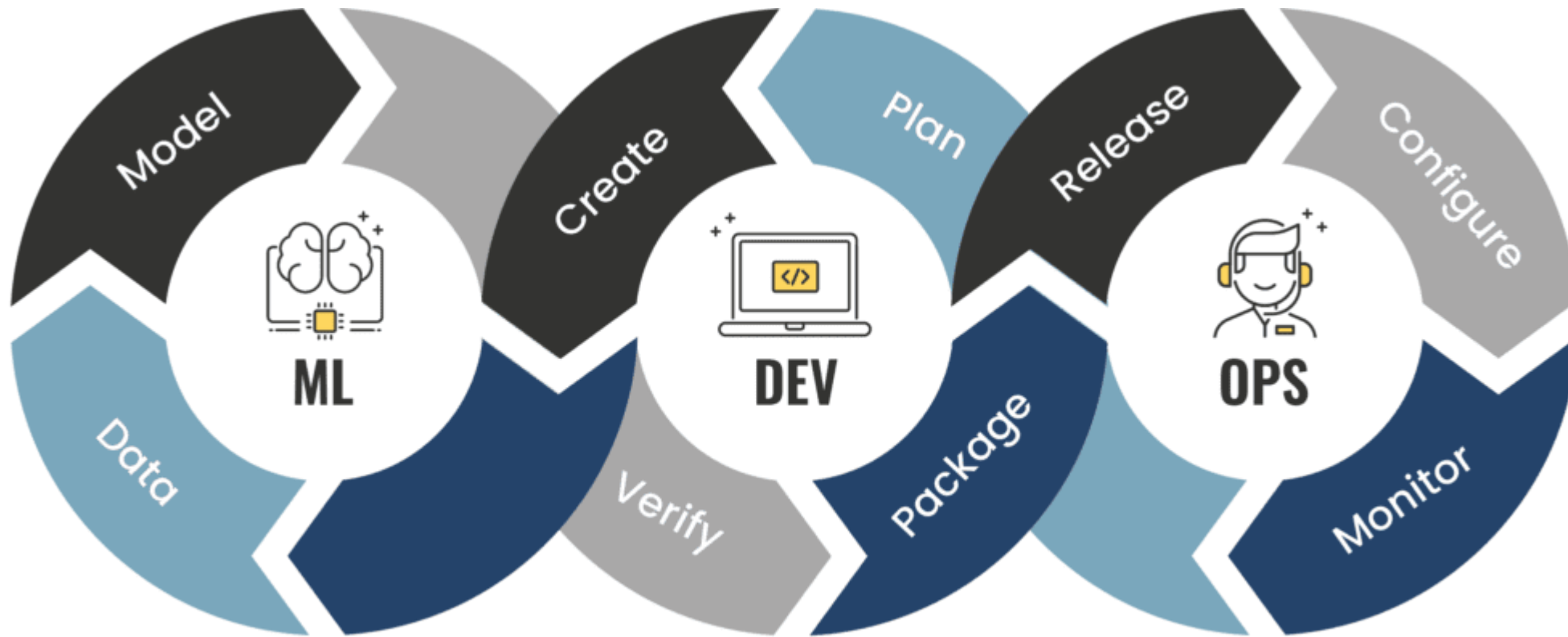
Om de gjør det, er det altfor mange som ikke blir implementert skikkelig.

Vi får mer oppmerksomhet rundt MLOps – fremfor ML-PoCs (2/3)

Det krever innsats for å skalere og integrere modellene til å kunne kjøre i produksjon, og vi må gå opp monitorering, vedlikehold og videreutvikling av dataflyt og modellflyt.

Heldigvis er det økende oppmerksomhet rundt MLOps, og det er en del virksomheter som er gode. Denne kunnskapen er i ferd med å spres, og **vi får mer ressurser rettet mot å få ting i produksjon** - skikkelig - fremfor å starte nok en PoC.

Vi får mer oppmerksomhet rundt MLOps – fremfor ML-PoCs (3/3)



<https://www.phdata.io/blog/mlops-vs-devops-whats-the-difference/>

5. Data science og analyse: Leve forretningsanalytikerne!

Forretningsanalytikere først (1/2)

Vi er ute etter å løse forretningsproblemer og bruke data til å støtte operativ drift.

Men - de fleste sliter med å bruke data til å svare på forretnings spørsmål som “hva har skjedd”/“hvorforskjedde det”.

For disse virksomhetene vil en **oppskalering av den kollektive datakompetansen ha mye større verdi** enn å optimalisere smale use case basert på AI.

Forretningsanalytikere først (2/2)

Derfor er spådommen dette: Data science som begrep roer seg ned litt, og **forretningsanalytikerne inntar rampelyset**.

Forretningsanalytikerne kan vi som regel dyrke frem selv, mens produktive data scientists er vanskeligere å hente inn. Så får vi heller bli mer avanserte etter hvert.

Økt bevissthet rundt kompetanseheving? Ja takk!

For mange ligger det fortsatt stort potensiale for verdiskapning gjennom å øke den felles evnen til å utnytte data.

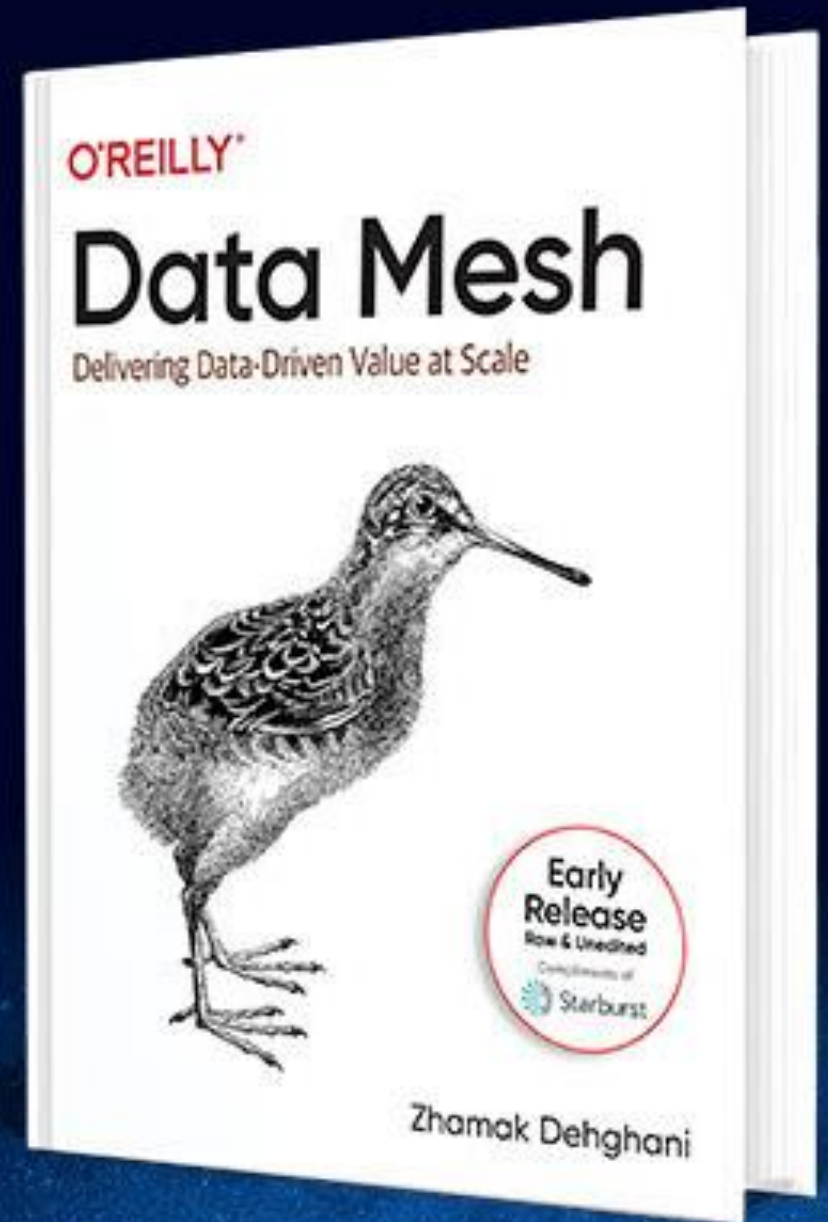
Interne kompetanseprogrammer fortsetter å bli gjennomført for å øke kompetansen, men dessverre vil fortsatt de fleste være fokusert på hvordan bruke Excel/Power BI/Tableau og andre selvbetjeningsverktøy, fremfor å bruke data til problemløsning.

6. Leveransemodell:
Desentralisert eierskap skyter
fart, men tilpasses virkeligheten

To mesh or not to mesh

Jobber du innen data-verdenen har du sikkert fått med deg begrepet data mesh.

Begrepet ble først beskrevet i to blogginnlegg skrevet av Zhamak Dehghani i 2019: "How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh" og "Data Mesh Principles and Logical Architecture".



Vi ønsker å unngå flaskehalsene

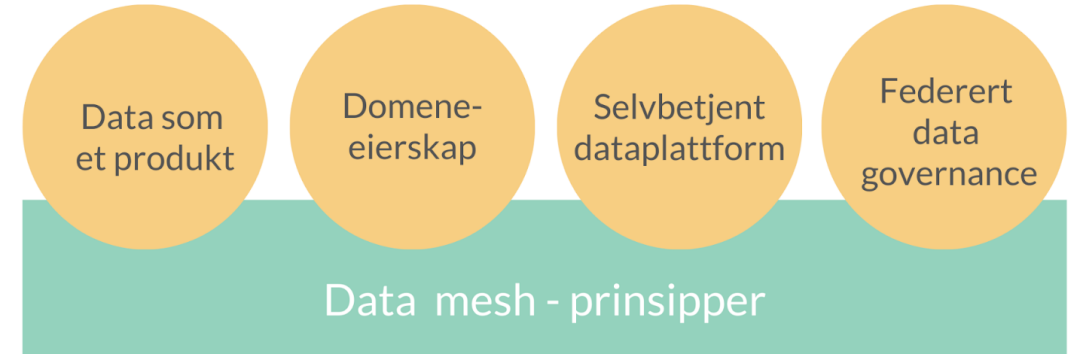
Dagens organisering skaper **flaskehals** gjennom et sentralt team av utviklere og analytikere, og en sentral dataplattform.

Data Mesh søker å fri seg fra en sentralisert arkitektur ved å desentralisere både eierskap og ansvar for å utvikle dataprodukter.



Produkter, Domener, Selvbetjening og Felles retningslinjer

Kjerneideen er at virksomheter kan bli mer datadrevet ved å flytte til domene-orientert dataeierskap og arkitektur, drevet av selvbetjent analyse og et sett felles retningslinjer.



Så til spådommen: Desentralisert eierskap skyter fart

Selskap som Zalando, ODA og Adevinta er nå aktive med å fortelle hvordan de implementerer data mesh.

"Ekte" data mesh har de færreste enda, siden de fleste organisasjoner må skalere opp analyse- og teknisk kompetanse betydelig.

Men – flere oppdager at data lever i en prosess, som noen eier

Mange er i ferd med å innse at det blir krevende å skille ut eierskaps- og leveransemodeller kun for dataprodukter.

Eierskapet til en prosess omfatter både mål og resultater, systemer, prosesser, organisering, kunder **OG data.**

Og: data mesh som designkonsept passer ikke alle typer virksomheter eller alle situasjoner

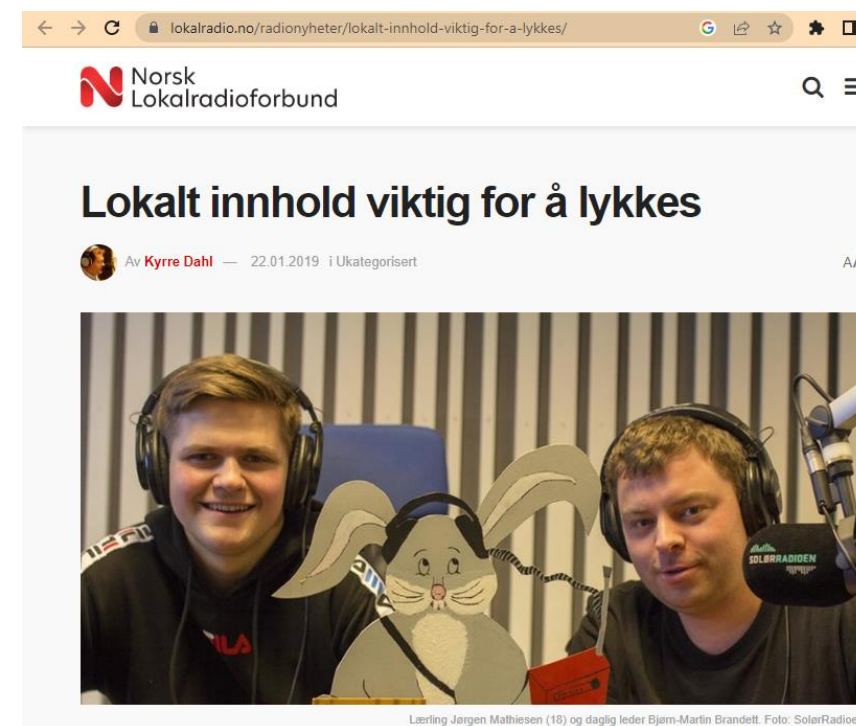
Desentralisert eierskap til datasett, rapporter og analysemodeller brer mer om seg, men vi er ikke fanatiske.

Det tar lengre tid med dataflyter og datadomener (tenk “Kunde” eller “Produkt” som domener).

Det kan også være verdt en debatt om man trenger å desentralisere arkitekturen

Vi får til mye med teknologier som Snowflake, Databricks og dbt i en sentral plattform, som logisk kan skilles ut i domener.

Så slutt å lage lokale dataplattformer.
Lag lokalt innhold.



7. Verdi fra data:
Alle innser at de må bli
datadrevet. Og AI til folket

ChatGPT åpner øynene til alle for verdien til AI (1/4)

ChatGPT har åpnet øynene til mange for verdien av data - og nå har over 100 millioner brukere testet hvordan såkalt generativ AI kan brukes til å automatisere oppgaver som tekstskriving og finne frem fakta.



ChatGPT åpner øynene til alle for verdien til AI (2/4)

Mange flere skal det bli i løpet av året, når løsningene som Microsoft Teams begynner å bygge inn kapabilitetene inn i tjenestene sine. For eksempel får vi nå møtereferat med aksjonspunkter servert rett etter møtet.

7. Verdi fra data: Alle innsær at de må bli datadrevet. Og AI til folket

The screenshot displays a Microsoft Teams meeting titled "Sales Analysis Review". The meeting is in progress, with a video grid showing participants: Serena Davis, Aadi Kapoor, Ray Tanaka, Danielle Booker, and Krystal. The interface includes a top navigation bar with "Sales Analysis Review", "Chat", "Files", "Details", and "Recap". A search bar is visible in the top right. An orange text box in the upper right corner contains the URL: <https://techmonitor.ai/technology/ai-and-automation/microsoft-to-integrate-chatgpt-into-teams>. The right sidebar features "Meeting content" with links to "Sales report Q4...", "Capacity stats list...", and "VanArsdelPitchDe...". Below this are tabs for "Notes", "AI Notes", "Mentions", "Transcript", and "Chat". A section titled "Your personal AI suggested notes and tasks" includes "Suggested notes" and "Suggested tasks".

Suggested notes

- ▶ **Serena** wants to look at the sales report before she and **Beth** spend more budget on the campaign. 5:00
- ▶ **Beth** explains that they are on track for new product release in December. But they will need to keep an eye on advertisement budget. She will follow up on that. 5:05
- ▶ **Beth** explains that they are managing the capacity well. They could be a problem if they get more popular. **Babak** is going to double check on that. 5:34

Suggested tasks

- ▶ Beth will follow up on advertisement budget.
- ▶ **Jon Shammass** will double check with **Amanda** regarding the ETA for the release of the new product.

ChatGPT åpner øynene til alle for verdien til AI (4/4)

Skoleelever og studenter verden over ser muligheten til å effektivisere oppgaveskrivingen - til stor glede.

Samtidig vil de oppdage at AI har bias, at tekst-algoritmer ikke kan regne og at fakta ikke alltid er fakta.

Kan vi begynne å snakke mer om fremtiden? (1/2)

Vi modner oss. Vi prater nå mindre om avansert analyse, og mer om å bli datadrevet gjennom alle prosesser.

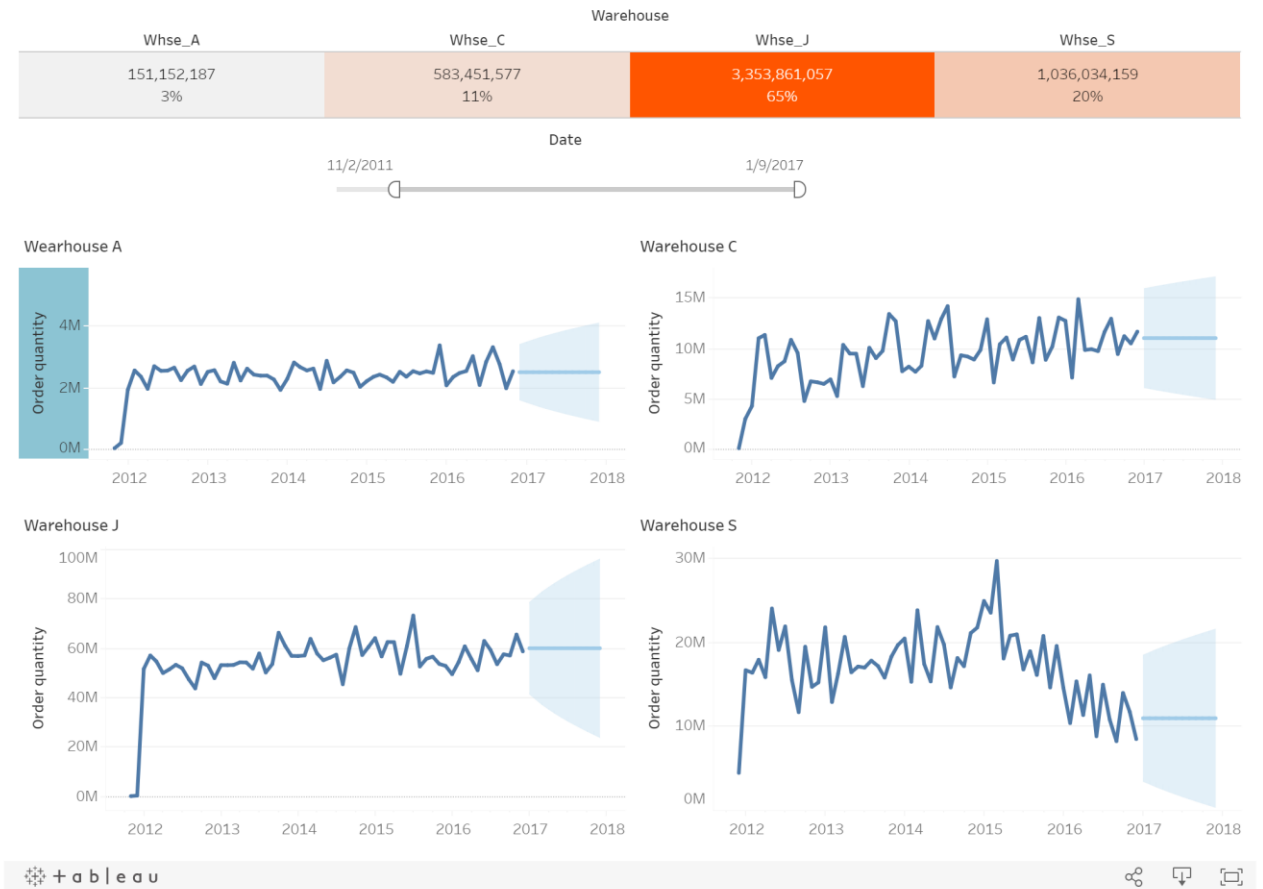
Vi ønsker å bruke data til å få faktabaserte beslutninger, kunne se mer fremover gjennom prediksjon og automatisere der vi har repetitive oppgaver som følger standardmønstre.

Kan vi begynne å snakke mer om fremtiden? (2/2)

Fortid er fortid. Det er fremtiden vi skal forberede oss på.

Jeg skulle ønske at flere ledere spurte forretningsanalytikerne sine om vi kan forsøke å se på **hva som kan skje fremover**: neste uke, neste måned og neste år.

Product demand forecasting



Prosessforbedringsarbeid blir mer datadrevet

Personlig har jeg et håp om at **prosessforbedring**, der Lean har vært moteordet i årtier, på sikt kan fusjonere med **datadrevet automatisering og problemløsning**.

De store konsulenthusene har nå både Lean-konsulenter og datakonsulenter, men samordningen og samarbeidet har ikke vært imponerende så langt. 2023 MÅ bli året der noen griper muligheten, og skaper de første norske suksesshistoriene som alle snakker om.

Glitni jobber med én ting: skape verdi fra data

Vi er i ferd med å samle de flinkeste konsulentene innen data og analyse, som brenner for å omsette innsikt og automatisering til verdi





5 ting å tenke på før dataplattformen står klar

5 ting å tenke på før dataplattformen står klar

📅 22.12.2022 | 🕒 5 min lesetid

Hvordan forbereder vi oss og de rundt oss på ny dataplattform? Samtidig som det bygges dataplattform bygges det også forventninger til hva som er neste steg. Her er 5 ting det er verdt å ha tenkt på før implementeringen er ferdig.



Glitni inngår partneravtaler

Glitni inngår partneravtaler

📅 15.12.2022 | 🕒 2 min lesetid

Glitni har nå inngått partnerskap med flere av de største navnene innen moderne dataplattformer: dbt Labs, Snowflake, Google Cloud og Databricks.



Hvorfor trenger du en datastrategi?

Hvorfor trenger du en datastrategi?

📅 29.11.2022 | 🕒 4 min lesetid

Er det nødvendig med en egen strategi for data? Les om hva en datastrategi skal hjelpe deg med og hvorfor du kanskje bør ta en ekstra titt på datastrategien du allerede har.



Vi slår et slag for å ha et sentralt dataplattformsteam

Vi slår et slag for å ha et sentralt dataplattformsteam

📅 25.10.2022 | 🕒 3 min lesetid

Skybasert arkitektur endrer hvordan vi jobber og organiserer oss. Her er grunnen til at du trenger et sentralisert plattformsteam.



Oppsummering av Coalesce 2022

Oppsummering av Coalesce 2022

📅 25.10.2022 | 🕒 3 min lesetid

Vi tar en rask titt på lanseringene til dbt Labs under Coalesce 2022, som inkluderte et nytt semantisk lag og støtte for python-utviklede modeller.



Hva er en dataplattform - og hvilke komponenter kan ligge til grunn?

Hva er en dataplattform - og hvilke komponenter kan ligge til grunn?

📅 28.08.2022 | 🕒 2 min lesetid

Du har kanskje endelig skjont forskjellen mellom et datavarehus og en data lake. Men så kom det enda et begrep: data lakehouse. Fortvil ikke, Glitni hjelper deg å navigere i begrepsjungelen!

Følg oss!

glitni.no
linkedin.com/company/glitni
medium.com/glitni

Snakk med oss!



Magne Bakkeli

Rådgiver / daglig leder

magne.bakkeli@glitni.no

91 66 22 69

glitni.no