



TOLLETATEN

Tolletaten, Power BI og Fabric

Erfaringer så langt

Prosjekt virksomhetsrapportering

Paul Osuldsen



Prosjektleder

- **Utdanning:** Siviløkonom
- **Erfaring:**
 - 30 års erfaring med virksomhetsstyring fra mer enn 100 norske og internasjonale bedrifter.
- **Power Platform:**
 - 30 års erfaring med BI verktøy og 8 års erfaring med Power BI

Ivar Handal



Prosjektmedarbeider (økonomi/virksomhetsstyring)

- **Utdanning:** siviløkonom med fordypning i finans
- **Erfaring:**
 - 19 års erfaring med virksomhetsstyring fra DFØ (6), Storebrand (10) og Tolletaten (3)
 - SAP, Unit4, Excel
- **Power Platform:**
 - 5 års erfaring med Power BI
 - 2 års erfaring med Power Apps og Power Automate













Øvrige team-medlemmer

+1 HR-ressurs (0,5 årsverk, i fødselspermisjon de neste 9 mnd)

+ 1 produkteier fra Styringsavdelingen (0,25 årsverk)

+ 1 innleid ressurs med Azure- og Fabric kompetanse (oktober – november 2023)

Innhold

  Tolletaten	Bakgrunn og formål for satsing på Power BI og dataanalyse i Tolletaten
   PDF	Tollinnsikt (tidligere «månedssrapporten») <ul style="list-style-type: none">• App og eksempler på rapporter
  	Arkitektur til nå («Ivar-modellen»/Citizen Developer-modell)
       	DFØ dataplattform <ul style="list-style-type: none">• Behov for ekstra Azure-tjenester for autentisering og data engineering (datainnhenting via API, relativt avanserte transformasjoner)• Fabric i bruk (vi er i test/utvikling) f.o.m. oktober 2023 → nye komponenter i arkitekturen vår• Fra Citizen Development til Citizen/Developer-modell
	Status erfaringer og veien videre

Bakgrunn og formål

- Uttrykt ønske fra **tolldirektør** om hyppigere og bedre styringsinformasjon
 - Dette behovet ble først imøtekommet med en fast månedsrapport i Word som ble manuelt oppdatert.
 - Rapporten i Word ble etter hvert erstattet med en Power BI-app med stort sett automatisk oppdaterte analyser/visualiseringer
- Tolletaten innførte i 2020 **M365 inkl. Power BI**. Dette beredet grunnen for å arbeide med styringsinformasjon på nye og mer effektive måter ved at data ble lettere tilgjengelig og lettere å dele internt i organisasjonen
 - Vi opplever **større og større etterspørsel** etter data-/informasjonsprodukter og er avhengige av å automatisere mest mulig/alt for å kunne imøtekomme etterspørselen
 - Regnskapsdataene fra DFØ er i dag 100% automatisert, mens vi baserer oss på **manuelle uttrekk** i Excel fra SAP/Innsikt



UNIT4

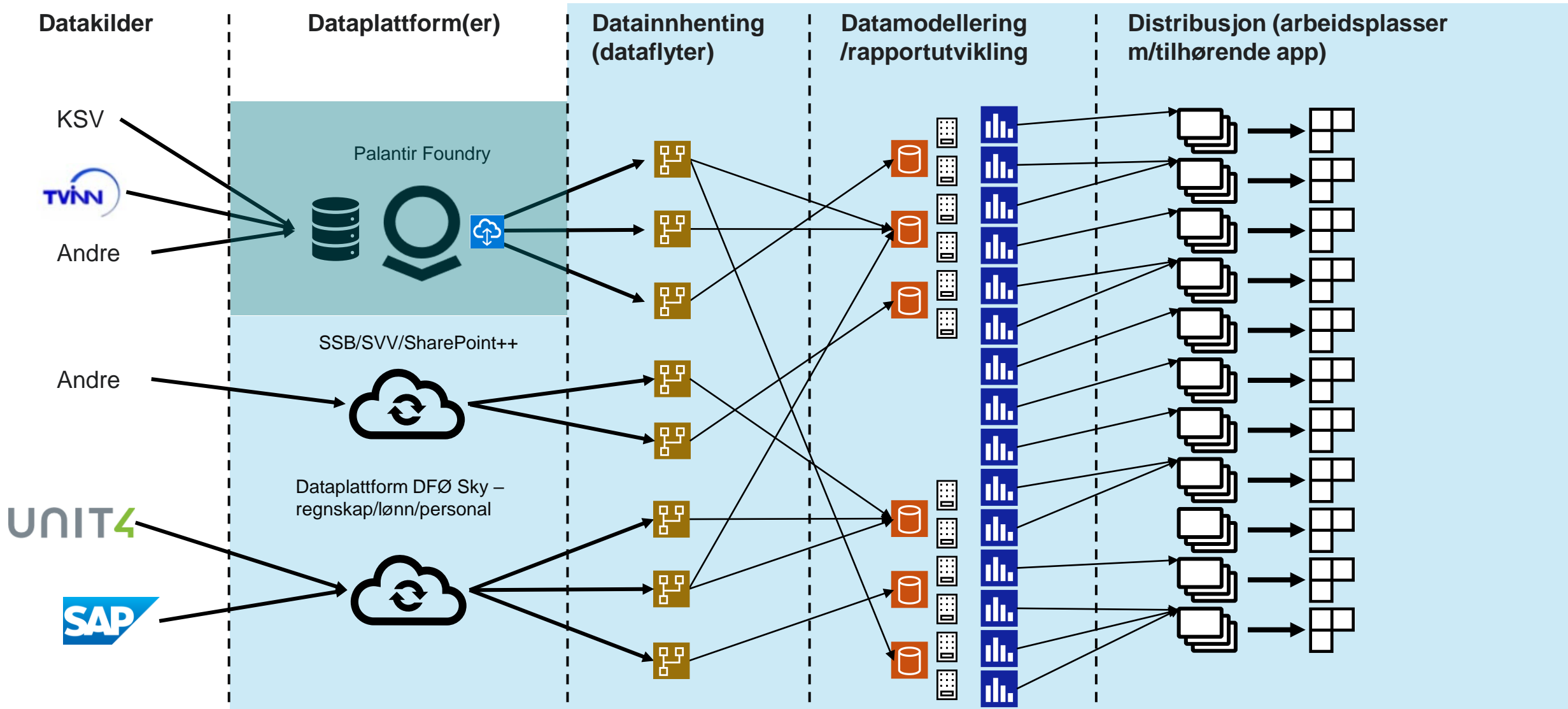
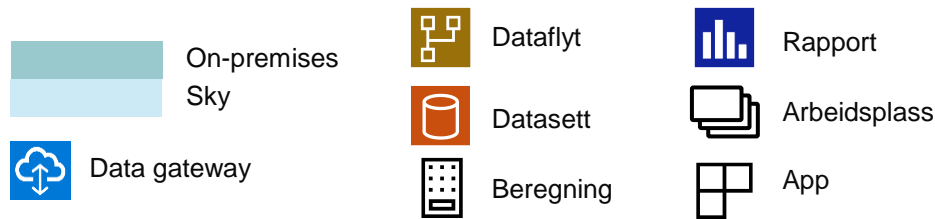


Historisk tilbakeblikk

- **Jan 2023:** Oppstart Prosjekt Virksomhetsrapportering. Arbeid med automatisering av Tolldirektørens månedsrapport (på papir/pdf). Målgruppe: ledere på nivå 1-3 (ca 65 brukere/lesere)
- **Mar 2023:** Tolldirektørens månedsrapport ble døpt om til «Tollinnsikt»
- **Apr 2023:** lansering av kjøreplan for Tollinnsikt
- **Oktober 2023:** vi tok i bruk Fabric-komponenter for å kunne hente inn HR-data fra DFØ dataplattform.
- **Jan 2023→nå:** det er utviklet en rekke løsninger (Økonomirapportering, Oppdragsstyring, Oversikt kontroll/beslag, Kjørebok/driftsjournal kjøretøyspark, Avviksregistrering m.m.) med målgruppe lik alle ledere på alle nivåer + ansatte i hele organisasjonen. Løsningene er utviklet i både Power BI, Power Apps og Dataverse.

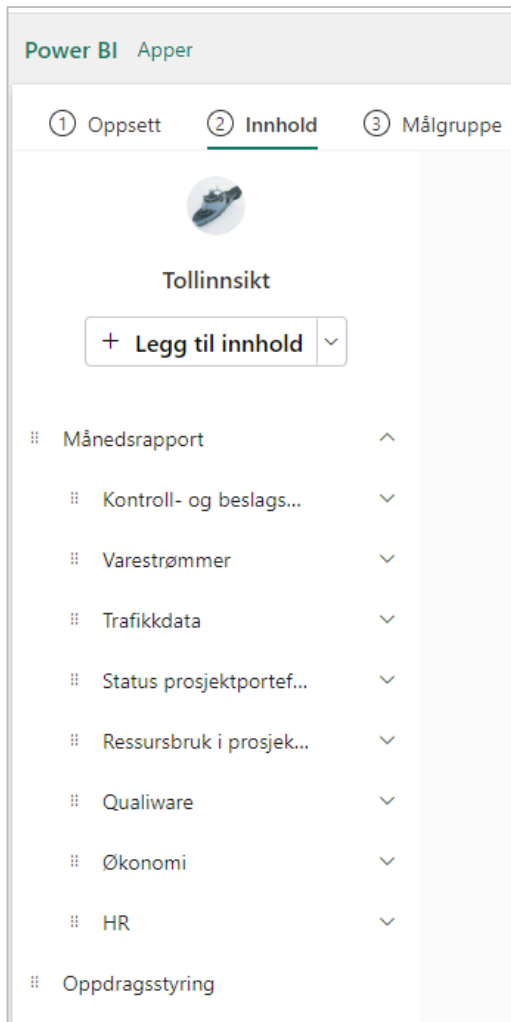


Arkitektur Prosjekt Virksomhetsrapportering - «Ivar»-modellen



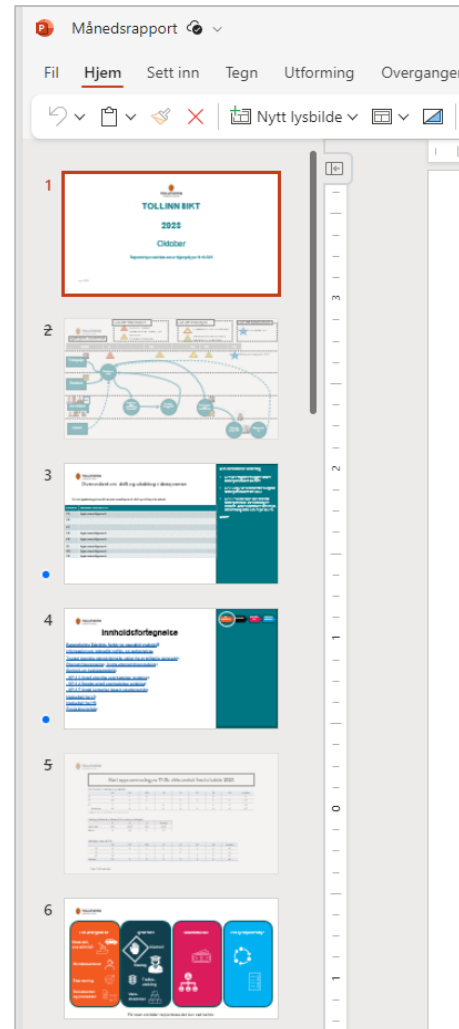
Tollinnsikt – en app, en PowerPoint og en PDF

En Power BI app



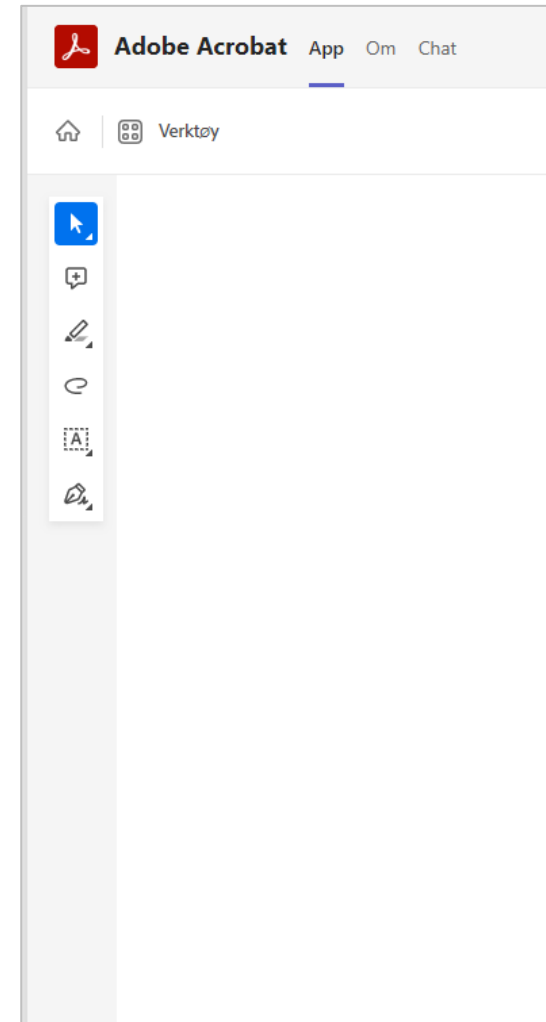
The screenshot shows the Power BI app interface. At the top, there are three tabs: "Oppsett", "Innhold", and "Målgruppe". The "Innhold" tab is selected. Below the tabs, there is a section for "Tollinnsikt" with a "+ Legg til innhold" button. A list of reports is displayed on the right side, including "Månedrapport", "Kontroll- og beslags...", "Varestrømmer", "Trafikkdata", "Status prosjektportef...", "Ressursbruk i prosjek...", "Qualiware", "Økonomi", "HR", and "Oppdragsstyring".

En PowerPoint



The screenshot shows a PowerPoint presentation titled "Månedrapport". The slide content includes a title slide "TOLLINN SIKT 2022 Oktober", a flowchart, a table, a list of items, and a summary slide with four colored boxes. The interface includes a menu bar with "Fil", "Hjem", "Sett inn", "Tegn", "Utforming", and "Overganger".

En PDF



The screenshot shows the Adobe Acrobat PDF viewer interface. The title bar reads "Adobe Acrobat App Om Chat". The main area is empty, and the left sidebar contains various tool icons for navigation and editing.



Kontroll- og beslagsstatistikk



Beslag

— Valgt år
- - - - - Benchmark

Tiltak

- Velg alt
- AOT
- Beslag
- Forelegg

Mål

- Antall
- Mengde
- Gj. snitt mengde

Periode

- Per periode
- Akkumulert

År

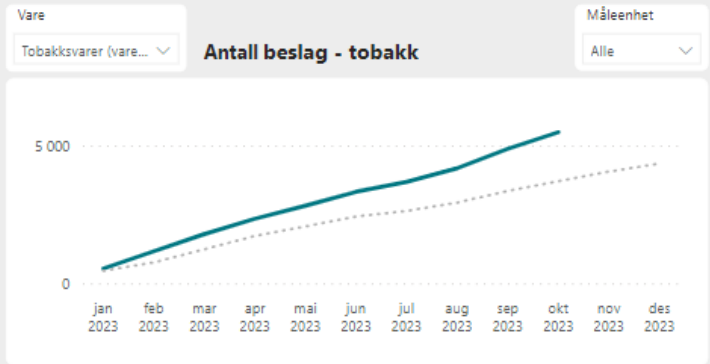
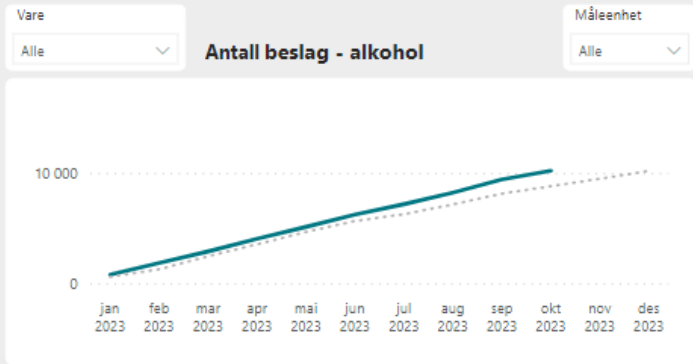
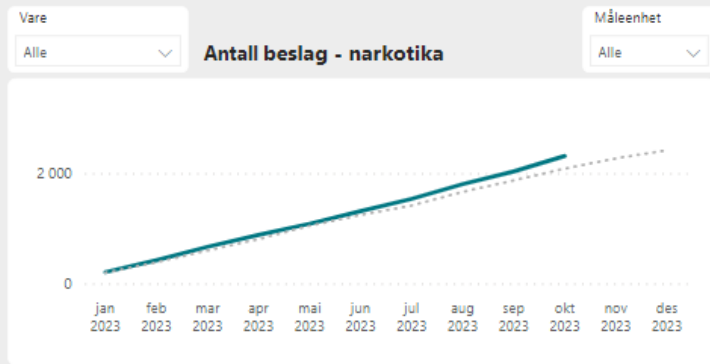
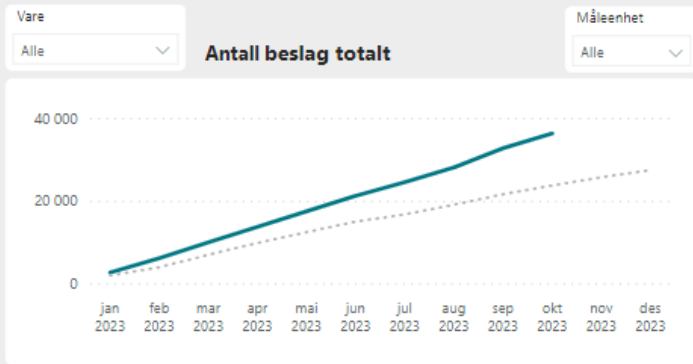
2023 2023

Benchmark

- 2019
- 2020
- 2021
- 2022
- 2022
- Ingen

Tid

- År
- Kvartal
- Måned



Vurdering

Her skal det skrives en vurdering av status og eventuelle tiltak.

Økonomistatus – post 01



TOLLETATEN STATUS BUDSJETT VS REGNSKAP - PER OKTOBER 2023

(tall i tusen kroner)



Budsjett

1 479 829

Regnskap

1 468 774

Avvik

11 054

Organisasjonsstruktur

Flere valg

Kontotype

Alle

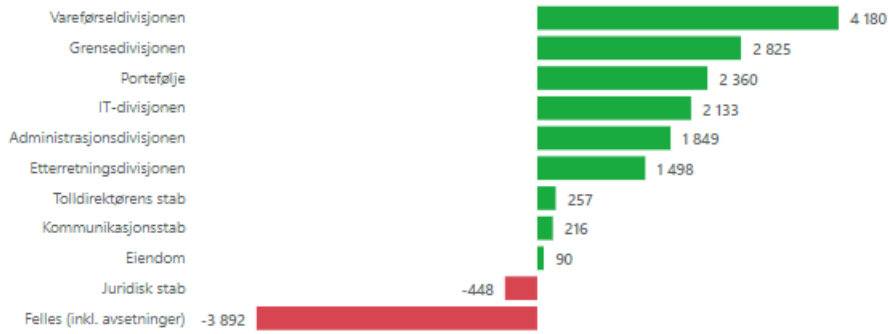
År og måned

Flere valg

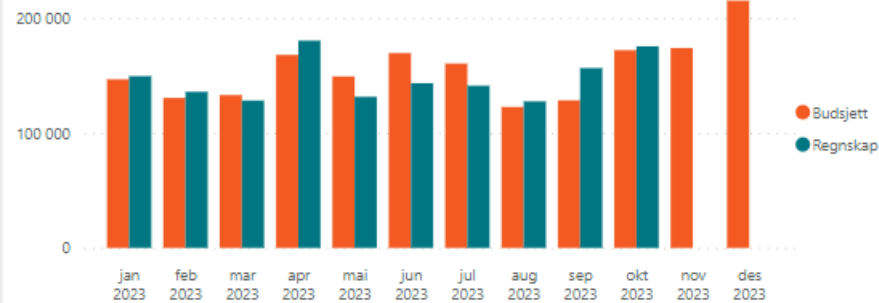
Finansiering

161001

Avvik hittil i år mellom regnskap og periodisert budsjett



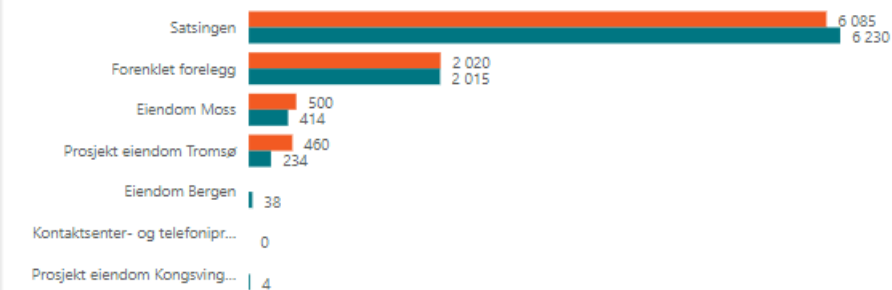
Måned for måned



Største avvik per artskonto



Porteføljeprosjekter - budsjett vs regnskap



Vurdering

Her skal det skrives en vurdering av status og eventuelle tiltak.

Oppdragsstyring - Tolletatens virksomhetsplan for perioden 2023-2026

Utviklingsgrep

Divisjon

🔍 Søk

- AD
- GD
- ID
- KS
- TS
- VD

Oppdragskategori

- Drift
- Utvikling

Prioritering

Alle

Virksomhetsstrategi

Utviklingsgrep

Oppdragskilde

Alle

Oppstart

Alle

Virksomhetsplan

2023-2026

ROD-visning

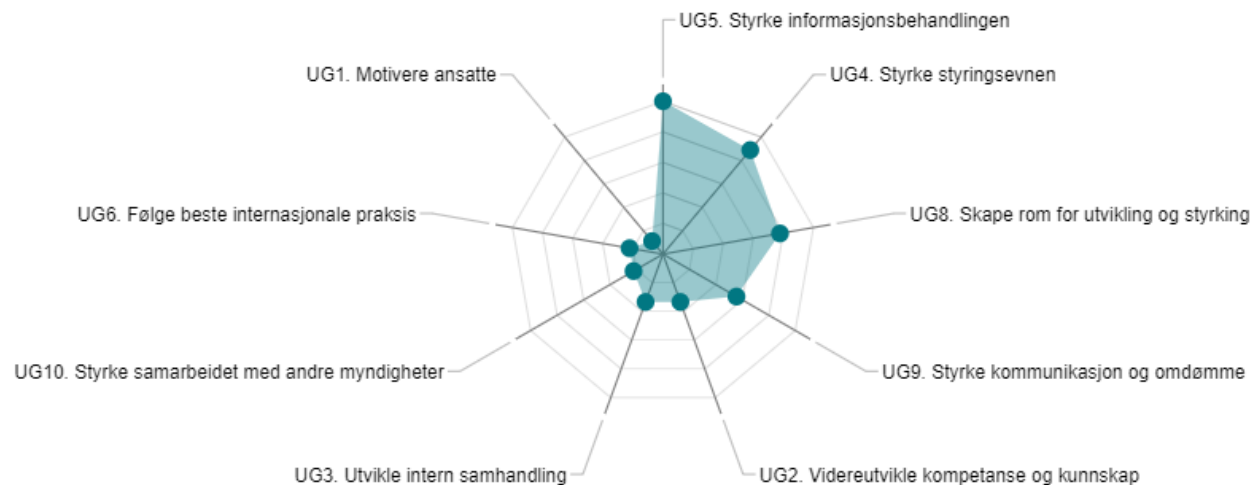
Oppdrag

Hovedgrep

Utviklingsgrep

Siste endring:

20.11.2023 11:41:00



Antall oppdrag

40

Herav oppstart 2021

1

Herav oppstart 2022

7

Herav oppstart 2023-->

32

Trykk på oppdrag

	Antall tiltak	Status	Prognose	Frist	Årsverk
<input type="checkbox"/> Innføre helhetlig styring og forvaltning av data og informasjon for Tolletaten	12	●	●	31.12.2026	8,8
<input type="checkbox"/> Fellesoppdrag AD - Effektivisering innenfor ADs fagområder	8	●	●	31.12.2026	3,0
<input type="checkbox"/> Fellesoppdrag Tolletaten - Dokumentere og vedlikeholde etatens prosesser	14	●	●	31.12.2026	3,0
<input type="checkbox"/> ID skal optimalisere verktøyene som understøtter egenutvikling i etaten for å sikre effektiv og moderne utviklingsprosesser	4	●	●		3,0
<input type="checkbox"/> ID skal sikre operasjonisering av operativt IT-sikkerhetsteam (CSIRT)	5	●	●		2,5
<input type="checkbox"/> ID skal sikre gjennomføring av tiltak besluttet i TLG etter revisjon av informasjonssikkerhet	5	●	●		2,0
<input type="checkbox"/> Videreutvikle og forvalte de tollfaglige utdanningsløpene - 2023	10	●	●		1,7
<input type="checkbox"/> ID skal etablere systemstøtte for virksomhetsrapportering basert på utredning	3	●	●		1,5
<input type="checkbox"/> AD skal videreutvikles anskaffelsesfunksjonen slik at viktige gevinster realiseres for frigjøring av ressurser til prioriterte aktiviteter	1	●	●	01.01.2024	1,0
<input type="checkbox"/> Forbedre mål- og resultatstyringen i etaten og utvikle gode styringsindikatorer	3	●	●	31.12.2026	1,0
<input type="checkbox"/> Implementere, drifte og utvikle prosessstyring og Qualiware	1	●	●	31.12.2026	1,0
<input type="checkbox"/> KS skal utforme, lansere og koordinere Tolletatens talspersonsprogram, herunder utvikling av budskapsnotater for prioriterte tematikker	1	●	●	29.12.2023	1,0
Totalt	132				37,7

Deling og samhandling rundt data/analyser vha bl.a. apper (Power Apps) i Teams

Oppdragsstyring Oppdragsstyring Om

Intern [Lær mer](#)

Status per nå: ● ● ● ● ●
Prognose per 31.12: ● ● ●

Avvik: Velg alt, Avvik, Ikke avvik

Trykk på oppdrag

Oppdrag	Antall tiltak	Status	Prognose
AD skal ivareta økonomiforvaltningen gjennom året på vegne av Tolldirektøren/TLG og levere økonomiske rapporter og regnskapsavleggelse til riktig tid og kvalitet til interessenter internt i etaten samt til FIN og Riksrevisjonen.	24	●	●
Ivareta forvaltning og utvikling av regelverk	16	●	●
ID skal levere drift og utvikling av etatens løsninger	15	●	●
Fellesoppdrag Tolletaten - Dokumentere og vedlikeholde etatens prosesser	14	●	●
Innføre helhetlig styring og forvaltning av data og informasjon for Tolletaten	12	●	●
Ivareta fagansvaret for virksomhetsstyring	12	●	●
Forberedelse, ledelse, deltakelse og annet arbeid med internasjonale møter	11	●	●
GD skal etablere struktur for, og gjennomføre utvikling- og vedlikehold av kompetanse for operativt tjenestepersonell	11	●	●
GD skal gjennomføre opplæringsløp for tollere uten tollfaglig utdanning som skal ha intensiv tollfaglig opplæring (ITO)	11	●	●
Videreutvikle og forvalte de tollfaglige utdanningsløpene - 2023	10	●	●
Virksomhetsstyring	10	●	●
Drive utviklingen av det internasjonale arbeidet i etaten	9	●	●
ED skal drifte og videreutvikle IT-støtte til etterretningsformål	9	●	●
Gjennomføre etatens planprosesser	9	●	●
Levere og rapportere styringsinformasjon på etatsnivå	9	●	●
ED skal levere kunnskap om vare- og trafikkstrømmer, trusler, aktører og handlemåter/moduser (data- og analysedrevet)	8	●	●
Fellesoppdrag AD - Effektivisering innenfor ADs fagområder	8	●	●
ID skal levere drift av etatens infrastruktur	8	●	●
Prosedyreseksjonen	8	●	●
Saksbehandling	8	●	●

Konsekvens og tiltak - status og prognose

Trykk på 1 oppdrag i tabellen til venstre

Divisjon: Søk

- AD
- ED
- GD
- ID
- JS
- KS
- TS
- VD

Oppdragskategori: Drift, Utvikling

Prioritering: Alle

Virksomhetsstrategi: Alle

Oppdragskilde: Alle

Oppstart: Alle

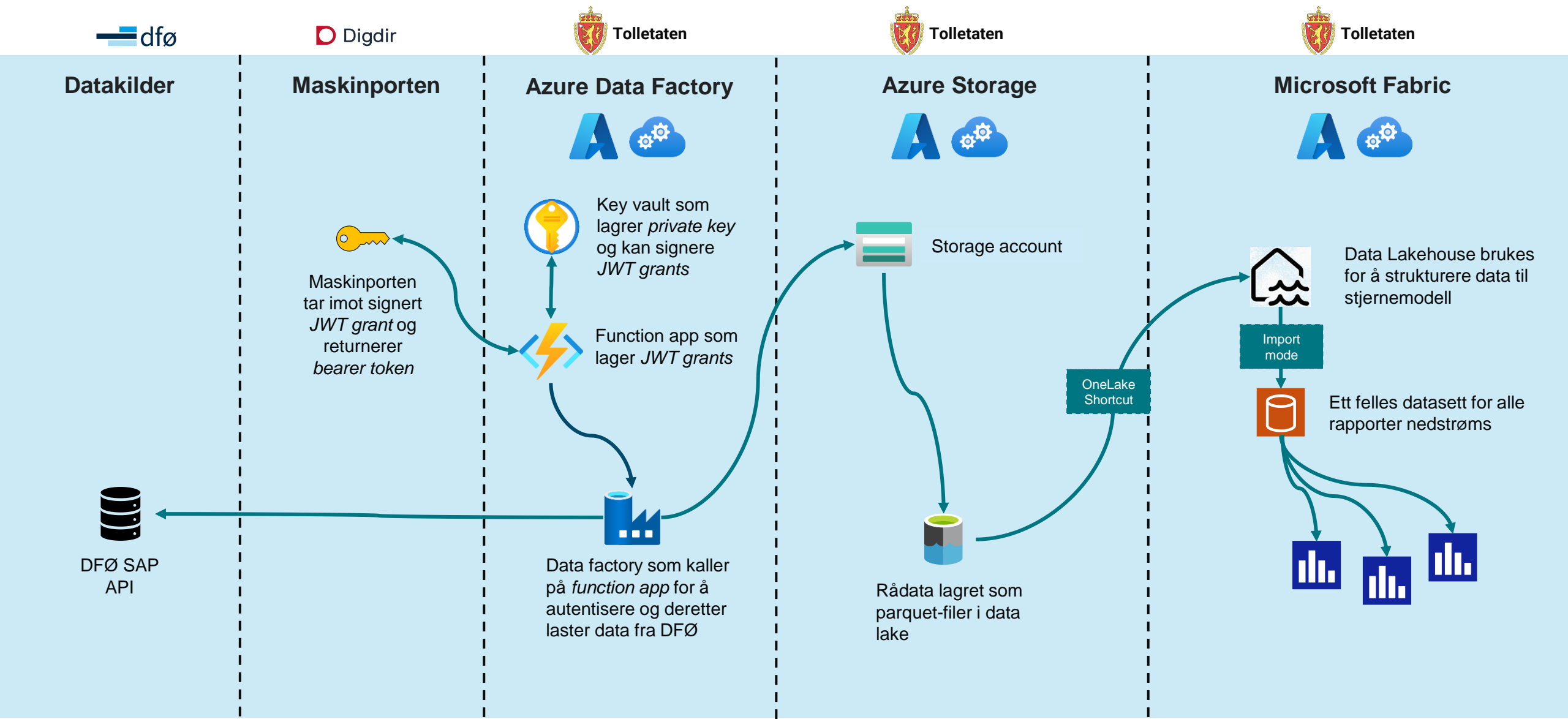
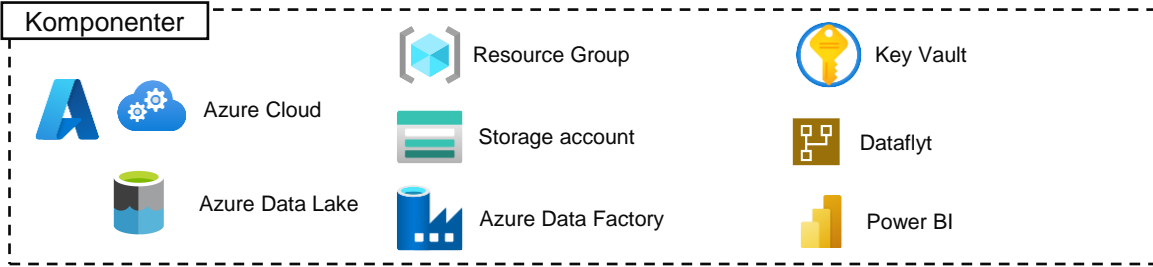
Virksomhetsplan: 2023-2026

ROD-visning: Oppdrag, Hovedgrep, Utviklingsgrep

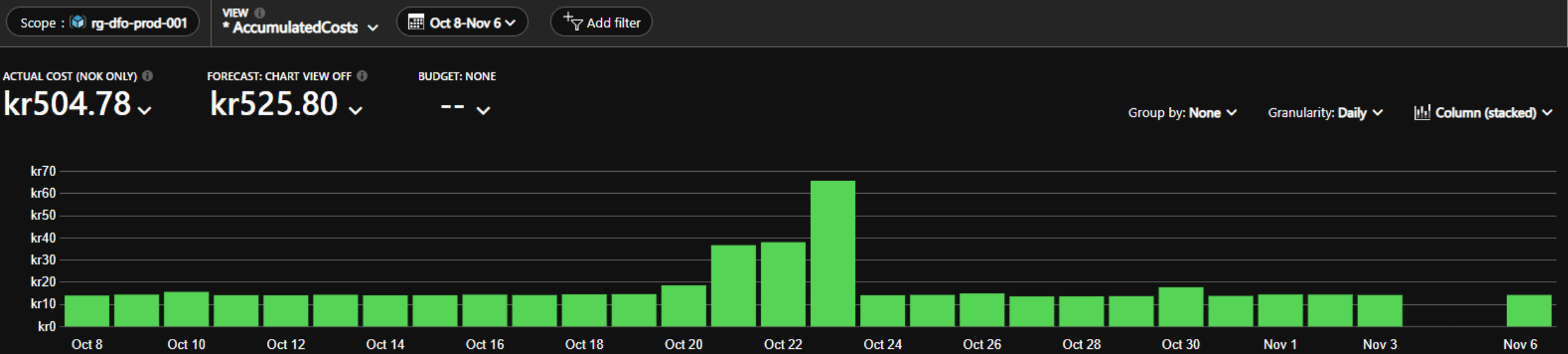
Siste endring:

DFØ dataplattform

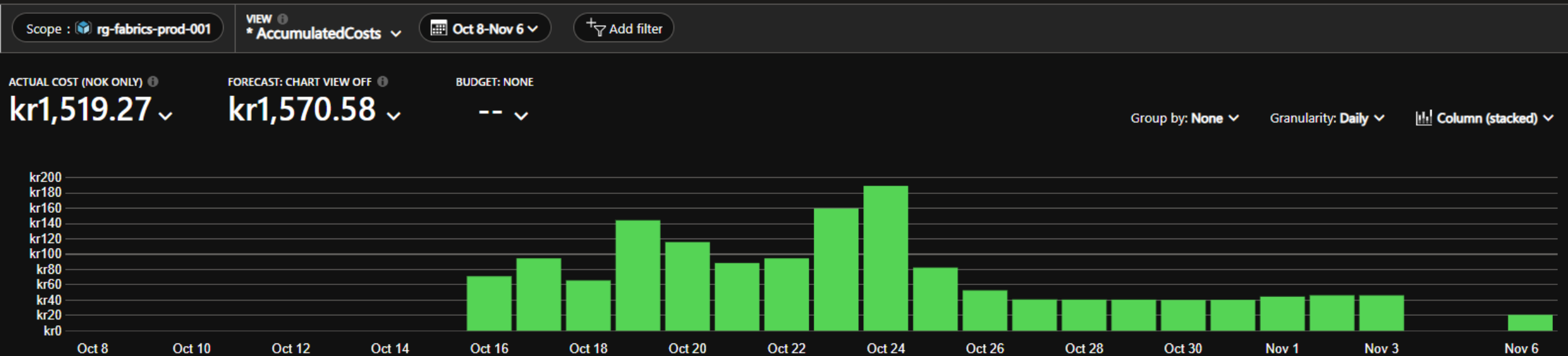
- Behov for ekstra Azure-tjenester for autentisering og data engineering mot maskinporten/API
 - Datainnhenting, relativt avanserte transformasjoner
- Fabric i bruk (vi er i test/utvikling) f.o.m. oktober 2023
- Nye komponenter i arkitekturen vår



Kostnader for Azure Data Factory, Function App, Key Store og Storage Account (data lake)



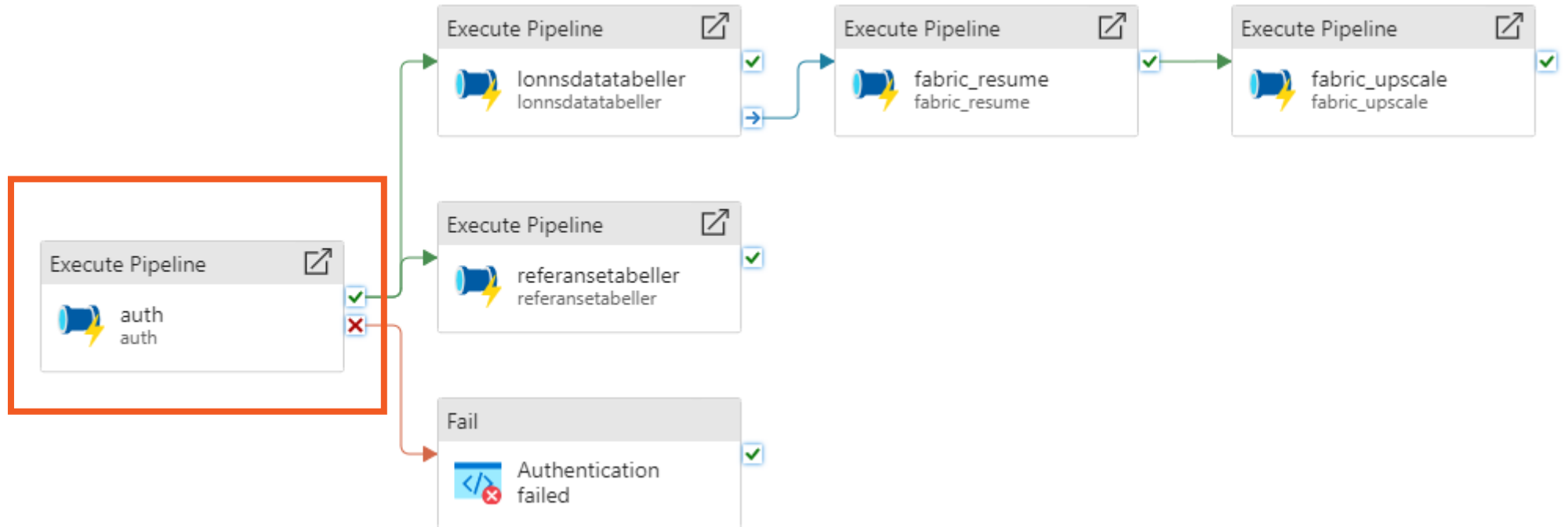
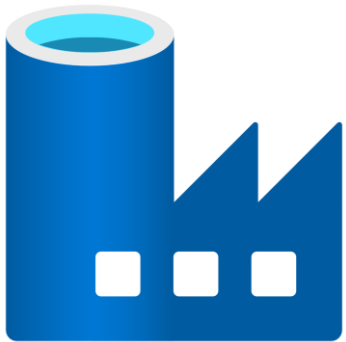
Kostnader for Microsoft Fabric



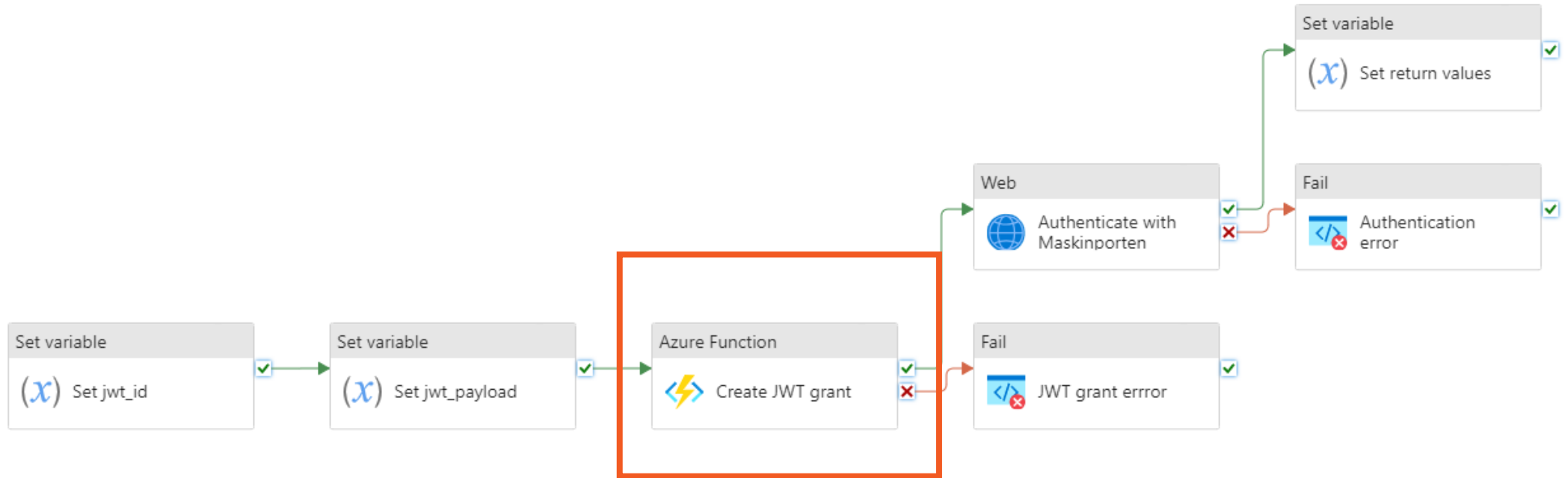
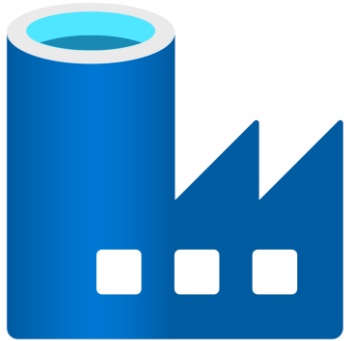
Status erfaringer og veien videre

- Data leses inn fra HR-API hver natt
- Mye jobb som krever spisskompetanse for å omforme/tilrettelegge data til analyseformål
 - API-et gir bare rådata fra SAP som krever mye tilrettelegging/transformasjon før det kan brukes i datamodeller/analyser
- Gjenstående arbeid er å teste data (avstemme mot SAP HR/Innsikt), optimalisere datamodeller og lage gode, brukervennlige rapporter
 - Årsrapport
 - Tertialrapport
 - Månedsrapport
 - Ad hoc-analyser
 - HR-administrasjon?

Azure Data Factory - master-pipeline som kjører kl. 02:30 hver natt



Azure Data Factory - autentiserings-pipeline



Azure Functions

Key Vault Azure Functions Maskinporten



+



+



Her er en kort forklaring på koden i fem nøkkelpunkter:

1. Oppsett av Azure-funksjon:

- Koden definerer en Azure-funksjon som utløses av en HTTP-forespørsel.
- Den setter standardverdier for nøkkelrelaterte parametere og henter inn data fra HTTP-forespørselen.

2. Integrasjon med Azure Key Vault:

- Koden kobler seg til Azure Key Vault ved hjelp av Azure-legitimasjon.
- Den henter en spesifisert nøkkel fra Key Vault og oppretter en kryptografiklient for signering.

3. Generering av JWT:

- Koden konstruerer overskriften og innholdet for en JSON Web Token (JWT).
- Den koder overskriften og innholdet, beregner en digest og signerer den med den private nøkkelen.

4. Tokenstruktur:

- Koden kombinerer den kodede tokenstrukturen og signaturen for å danne det endelige signerte JWT.
- Den bruker Base64-koding og URL-sikre tegn for å lage tokenet.

5. HTTP-respons:

- Koden returnerer det signerte tokenet som en HTTP-respons med statuskode 200.
- Den genererte JWT kan brukes til autentisering, spesielt i Azure-tjenestemiljøer.

```
# Connect to the key vault
credential = DefaultAzureCredential()
key_client = KeyClient(vault_url="https://kv-dfo.vault.azure.net/", credential=credential)

# Instantiate the private key we are going to sign the token with
key = key_client.get_key(key_name)
crypto_client = CryptographyClient(key, credential=credential)

jwt_iat = int(time.time()) # Issued at current time
jwt_exp = jwt_iat + 60    # Expires in one minute

# Header and payload (claims)
header = {
    "alg": "RS256",          # RS256, RS384 and RS512 are supported by Maskinporten.
    "kid": jwt_kid         # Key identifier the asymmetric key bound to the client.
}

payload = {
    "aud": 'https://maskinporten.no/', # Maskinporten endpoint
    "iss": jwt_iss,                   # Our Oauth client ID
    "iat": jwt_iat,                   # Issued at unix seconds
    "exp": jwt_exp,                   # Expires at unix seconds
    "jti": jwt_id                     # Our internal token identifier (optional)
}

# This allows us to return tokens with or without scopes
if jwt_scope:
    payload["scope"] = jwt_scope

# Base64url encode the JSON-encoded header and payload
encoded_header = base64.urlsafe_b64encode(json.dumps(header).encode()).decode().rstrip("=")
encoded_payload = base64.urlsafe_b64encode(json.dumps(payload).encode()).decode().rstrip("=")

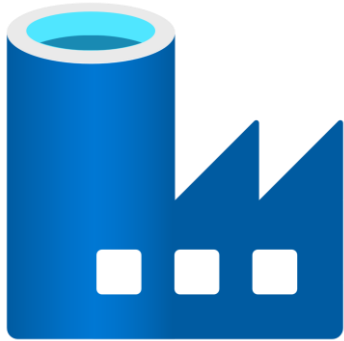
# Concatenate the encoded header and encoded payload with a period ('.') separator
token_body = f"{encoded_header}.{encoded_payload}"
digest = hashlib.sha256(token_body.encode()).digest()

# Sign returns the signature and the metadata required to verify it
result = crypto_client.sign(SignatureAlgorithm.rs256, digest)
token_signature = base64.urlsafe_b64encode(result.signature).decode('utf-8').rstrip("=")

# Concatenate the token body and the signature with a period ('.') separator
signed_token = f"{token_body}.{token_signature}"

return func.HttpResponse(
    signed_token,
    status_code=200
)
```

Azure Data Factory - autentiserings-pipeline



lonnsdatatabeller

Validate Debug Add trigger

Activities

Search activities

- Move and transform
- Synapse
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

Execute Pipeline

PA0000
PA0000

✖ ✓

🗑️ </> 📄 ➕

Execute Pipeline

PA0001
PA0001

Execute Pipeline

PA0002
PA0002

Execute Pipeline

PA0007
PA0007

Execute Pipeline

General **Settings** User properties

Invoked pipeline * PA0000 [Open](#) [+ New](#)

Wait on completion

Name	Type	Value	Default value
bearer_token	string	@pipeline().parameters.bearer_token	
table_name	string	PA0000	PA0000
scope	string	@pipeline().parameters.scope	lonnsdatatabeller
orderby	string	Value	BEGDA
query_fields	string	Value	PERNR,BEGDA,ENDDA,AEDTM,STAT2,STAT3,MASSN,MASSG

Azure Data Lake



Microsoft Azure

Home > rg-dfo-prod-001 > sadfo | Containers >

lonnsdatatabeller

Container

Search

Upload Add Directory Refresh Rename Delete Change tier

Authentication method: Access key ([Switch to Microsoft Entra user account](#))
Location: lonnsdatatabeller

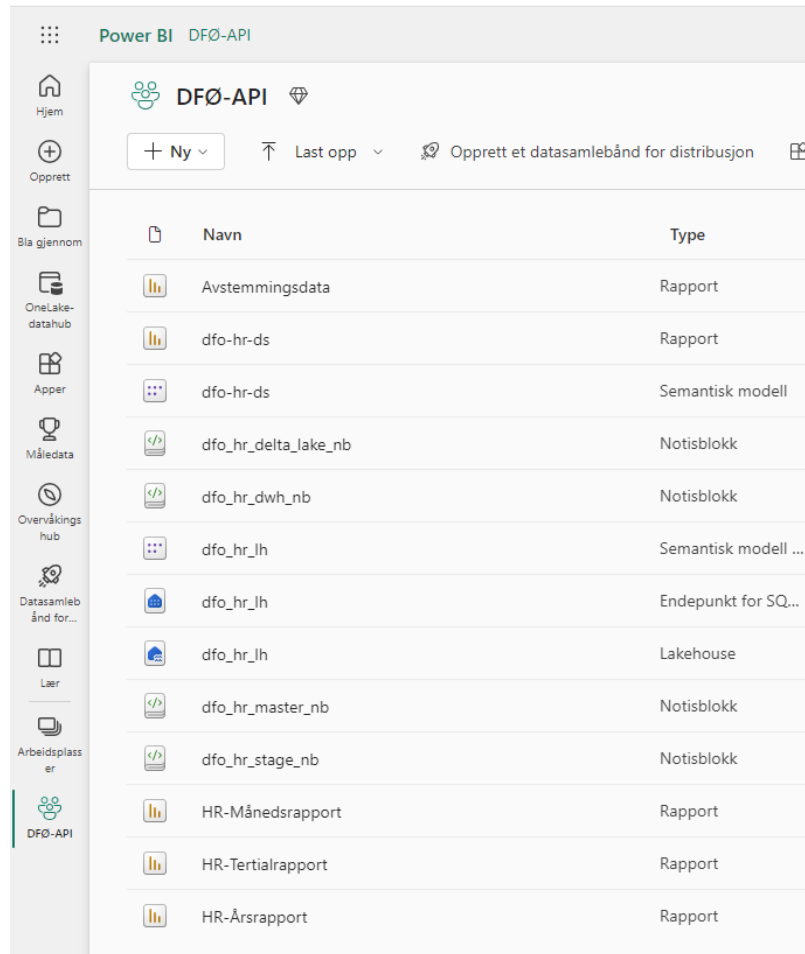
Search blobs by prefix (case-sensitive)

Name	Modified
<input type="checkbox"/> HRP1000	
<input type="checkbox"/> HRP1001	
<input type="checkbox"/> HRP1051	
<input type="checkbox"/> PA0000	
<input type="checkbox"/> PA0001	
<input type="checkbox"/> PA0002	
<input type="checkbox"/> PA0007	
<input type="checkbox"/> PA0008	
<input type="checkbox"/> PA0014	
<input type="checkbox"/> PA0022	
<input type="checkbox"/> PA0041	
<input type="checkbox"/> PA0509	
<input type="checkbox"/> PA2001	
<input type="checkbox"/> PA2002	

Settings

- Shared access tokens
- Manage ACL
- Access policy
- Properties
- Metadata

Fabric-elementer som vi jobber med



The screenshot shows the Power BI interface for a workspace named 'DFØ-API'. The left sidebar contains navigation icons for Home, Create, Browse, OneLake datahub, Apps, Measured data, Monitoring hub, Data connections, Learn, and Workspaces. The main area displays a table of Fabric elements:

Navn	Type
Avstemningsdata	Rapport
dfo-hr-ds	Rapport
dfo-hr-ds	Semantisk modell
dfo_hr_delta_lake_nb	Notisblokk
dfo_hr_dwh_nb	Notisblokk
dfo_hr_lh	Semantisk modell ...
dfo_hr_lh	Endepunkt for SQ...
dfo_hr_lh	Lakehouse
dfo_hr_master_nb	Notisblokk
dfo_hr_stage_nb	Notisblokk
HR-Månedssrapport	Rapport
HR-Tertialrapport	Rapport
HR-Årsrapport	Rapport

Power BI-elementer

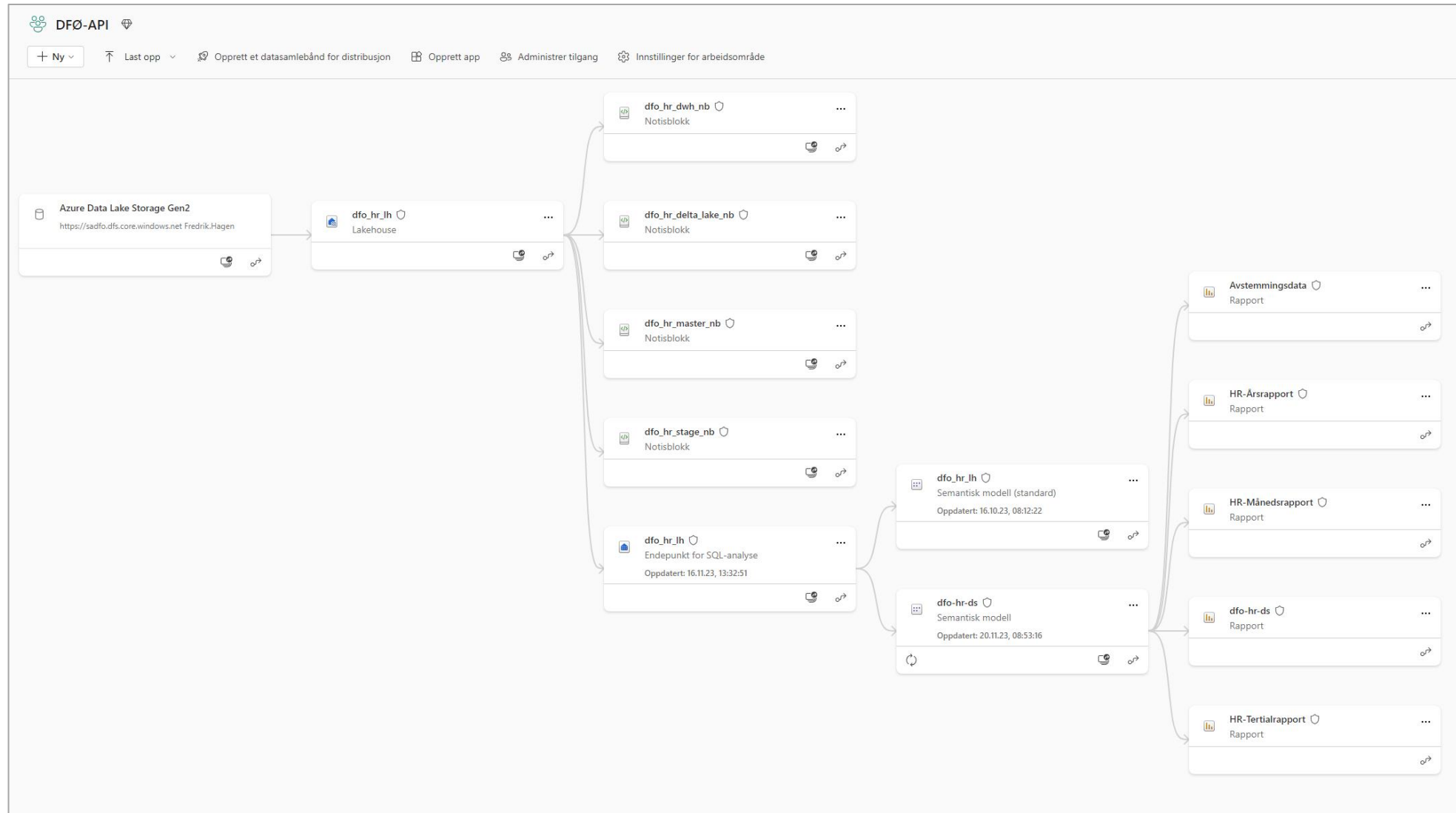
- Rapporter
- Semantisk modell (tidligere datasett)

Nye elementer fra Fabric

- Notebooks
 - Lar oss kjøre pySpark for å gjøre enkelte operasjoner/transformasjoner
- Lakehouse
 - «Fleksibelt datavarehus»
- Endepunkt for SQL
 - SQL-representasjonen av lakehouse. Kilde til Power BI. Abstraksjonslag for lakehouse.

Vi bruker p.t. **ikke direct lake** da vi er koblet til SQL-endepunkt og er på en billig versjon av Fabric (F2/F4), noe som ikke gir tilstrekkelig ytelse → import mode fra semantisk modell/datasett

Dataavstamming – hvordan flyter dataene?



Notebooks som kjøres hver natt

I første steget laster vi rådata fra parquet filene i vår data lake gjennom en shortcut og lagrer som delta. Her fåregår ingen transformeringer.

```
[1] 1 %run dfo_hr_delta_lake_nb
✓ -Command executed PySpark (Python) ▾
```

Det andre steget benyttes i hovedsak for å omstrukturere råtabellene fra periode med fra- og til-datoer, til å ha én rad per dato. Dette gjør det for eksempel mulig at sykefravær som går over nyttår kan telles i begge årene. Her ligger det det også annen forretningslogikk som er beskrevet mer detaljert i notebooken.

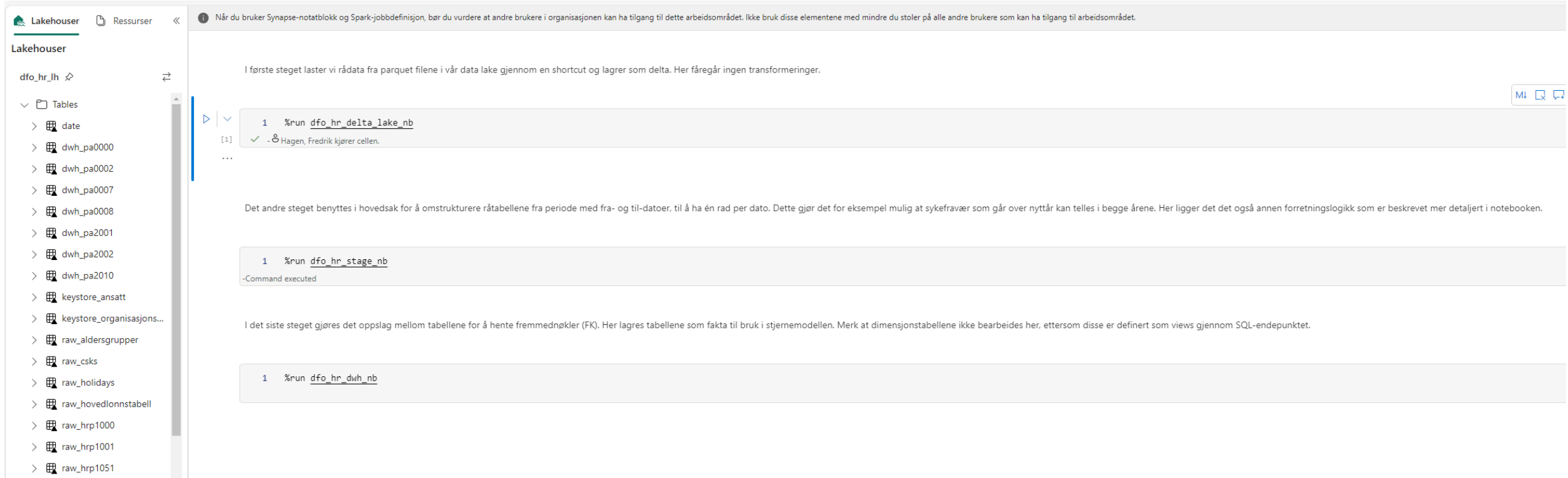
```
1 %run dfo_hr_stage_nb
-Command executed PySpark (Python) ▾
```

</> ✎ 🔒 ⋮ 🗑

I det siste steget gjøres det oppslag mellom tabellene for å hente fremmednøkler (FK). Her lagres tabellene som fakta til bruk i stjernemodellen. Merk at dimensjonstabellene ikke bearbeides her, ettersom disse er definert som views gjennom SQL-endepunktet.

```
1 %run dfo_hr_dwh_nb
PySpark (Python) ▾
```

Notebook 1- dfo_hr_master_nb



The screenshot shows a Lakehouse notebook interface. On the left, a sidebar lists tables under the folder 'dfo_hr_lh'. The main area contains three stages of a pipeline:

- Stage 1:** Command: `%run dfo_hr_delta_lake_nb`. Status: [1] ✓. Comment: Hagen, Fredrik kjører cellen.
- Stage 2:** Command: `%run dfo_hr_stage_nb`. Status: -Command executed.
- Stage 3:** Command: `%run dfo_hr_dwh_nb`.

Text descriptions for each stage:

- Stage 1:** I første steget laster vi rådata fra parquet filene i vår data lake gjennom en shortcut og lagrer som delta. Her fåregår ingen transformeringer.
- Stage 2:** Det andre steget benyttes i hovedsak for å omstrukturere råtabellene fra periode med fra- og til-datoer, til å ha én rad per dato. Dette gjør det for eksempel mulig at sykefravær som går over nyttår kan telles i begge årene. Her ligger det det også annen forretningslogikk som er beskrevet mer detaljert i notebooken.
- Stage 3:** I det siste steget gjøres det oppslag mellom tabellene for å hente fremmednøkler (FK). Her lagres tabellene som fakta til bruk i stjernemodellen. Merk at dimensjonstabellene ikke bearbeides her, ettersom disse er definert som views gjennom SQL-endepunktet.

Masterpipeline som kjøres hver natt og triggerer de andre notebookene



Notebook 2 - dfo_hr_delta_lake_nb

Fra filformat Parquet til Delta Parquet

• Last av PA-tabeller

Hver kodeblokk kjører en full last av PA-tabellene fra data lake (parquet) til delta lake. Dette steget er såpass kjapt og lettvent at det ikke er et stort behov for å etablere inkrementell last.

```
1 df = spark.read.load('Files/sadfo_lonnsdatatabeller/PA0000', format='parquet', header=True)
2 df.write.format("delta").mode("overwrite").option("overwriteSchema", "True").saveAsTable("raw_PA0000")
```

[1] ✓ -Command executed in 26 sek 5 ms by Hagen, Fredrik on 8:36:50 AM, 10/25/23

```
1 df = spark.read.load('Files/sadfo_lonnsdatatabeller/PA0001', format='parquet', header=True)
2 df.write.format("delta").mode("overwrite").option("overwriteSchema", "True").saveAsTable("raw_PA0001")
```

[1] ✓ -Command executed in 25 sek 739 ms by Hagen, Fredrik on 9:43:39 AM, 10/23/23

```
1 df = spark.read.load('Files/sadfo_lonnsdatatabeller/PA0002', format='parquet', header=True)
2 df.write.format("delta").mode("overwrite").option("overwriteSchema", "True").saveAsTable("raw_PA0002")
```

[2] ✓ -Command executed in 6 sek 422 ms by Hagen, Fredrik on 9:43:46 AM, 10/23/23

```
1 df = spark.read.load('Files/sadfo_lonnsdatatabeller/PA0007', format='parquet', header=True)
2 df.write.format("delta").mode("overwrite").option("overwriteSchema", "True").saveAsTable("raw_PA0007")
```

[3] ✓ -Command executed in 6 sek 552 ms by Hagen, Fredrik on 9:43:53 AM, 10/23/23

```
1 df = spark.read.load('Files/sadfo_lonnsdatatabeller/PA0008', format='parquet', header=True)
2 df.write.format("delta").mode("overwrite").option("overwriteSchema", "True").saveAsTable("raw_PA0008")
```

[4] ✓ -Command executed in 8 sek 191 ms by Hagen, Fredrik on 9:44:01 AM, 10/23/23

Notebook 3 - dfo_hr_stage_nb

Transformering/
tilrettelegging av
dataene til videre
arbeid/
analyseformål

Lakehouseer Ressurser <<

Lakehouseer

dfo_hr_lh ↗

Tables

- date
- dwh_pa0000
- dwh_pa0002
- dwh_pa0007
- dwh_pa0008
- dwh_pa2001
- dwh_pa2002
- dwh_pa2010
- keystore_ansatt
- keystore_organisasjons...
- raw_aldersgrupper
- raw_csks
- raw_holidays
- raw_hovedlonnstabell
- raw_hrp1000
- raw_hrp1001
- raw_hrp1051
- raw_pa0000
- raw_pa0001
- raw_pa0002
- raw_pa0007
- raw_pa0008
- raw_pa0041
- raw_pa0509
- raw_pa2001

Når du bruker Synapse-notatblokk og Spark-jobbdefinisjon, bør du vurdere at andre brukere i organisasjonen kan ha tilgang til dette arbeidsområdet. Ikke bruk disse elementene med mindre du stoler på alle andre brukere som kan ha tilgang til arbeidsområdet.

dfo_hr_stage_nb

Hver av kodeblokkene nedenfor joiner en PA-tabell med en datodimensjon og returnerer en rad per dato i tidsintervallet mellom BEGDA og ENDDA. Det betyr at alle tabellene får én rad per dag per ansatt innenfor gyldighetsintervallet.

Når det skal legges til nye tabeller er det viktig å huske:

1. BEGDA og ENDDA er integers, ikke datofelt. De må landes som integers i data laken for å ikke få datatype mismatch.
2. Ettersom BEGDDA og ENDDA er integers, må disse joines på NumericDate, ikke Date.
3. Dersom den nye tabellen inneholder datoer som faller utenfor datodimensjonen, må datodimensjonen utvides for å inneholde en komplett tidsserie.

```
1 df = spark.sql("""
2     SELECT
3         d.Date, -- Dato fra datotabell
4         p.OTYPE,
5         p.OBJID,
6         p.SHORT,
7         p.STEXT,
8         p.BEGDA,
9         p.ENDDA
10    FROM dfo_hr_lh.date d -- Generisk datotabell
11    INNER JOIN dfo_hr_lh.raw_hrp1000 p -- Valgt tabell
12        ON d.NumericDate >= p.BEGDA -- Dato større eller like startdato i SAP
13        AND d.NumericDate <= p.ENDDA -- Dato mindre eller lik slutt dato i SAP
14    """)
15
16 # Til slutt lagres den transformerte tabellen som delta
17 df.write.format("delta").mode("overwrite").option("overwriteSchema", "True").saveAsTable("stg_hrp1000")
```

```
1 df = spark.sql("""
2     SELECT d.Date, p.*
3     FROM dfo_hr_lh.date d
4     INNER JOIN dfo_hr_lh.raw_hrp1001 p
5         ON d.NumericDate >= p.BEGDA
6         AND d.NumericDate <= p.ENDDA
7     """)
8
9 df.write.format("delta").mode("overwrite").option("overwriteSchema", "True").saveAsTable("stg_hrp1001")
```

[2] ✓ -Command executed in 5 min 31 sek 441 ms by Hagen, Fredrik on 3:09:22 PM, 10/30/23

Notebook 4 - dfo_hr_dwh_nb

Bygging av faktatabeller

Lakehouser Ressurser

Lakehouser

- dfo_hr_lh
 - Tables
 - date
 - dwh_pa0000
 - dwh_pa0002
 - dwh_pa0007
 - dwh_pa0008
 - dwh_pa2001
 - dwh_pa2002
 - dwh_pa2010
 - keystore_ansatt
 - keystore_organisasjons...
 - raw_aldersgrupper
 - raw_csks
 - raw_holidays
 - raw_hovedlonnstabell
 - raw_hrp1000
 - raw_hrp1001
 - raw_hrp1051
 - raw_pa0000
 - raw_pa0001
 - raw_pa0002
 - raw_pa0007
 - raw_pa0008
 - raw_pa0041
 - raw_pa0509
 - raw_pa2001
 - raw_pa2002
 - raw_pa2010
 - raw_t001p

Når du bruker Synapse-notatblokk og Spark-jobbdefinisjon, bør du vurdere at andre brukere i organisasjonen kan ha tilgang til dette arbeidsområdet. Ikke bruk disse elementene med mindre du stoler på alle andre brukere som kan ha tilgang til arbeidsområdet.

Faktatabeller

I dette steget henter vi dimensjonsnøkler fra PA0001 (Organisasjonstilhørighet) og PA0002 (Ansatt), samt andre referansetabeller, og bygger faktatabeller.

PA0000 Personalendringer

Fakt_Personalendringer brukes per i dag kun til å kalkulere turnover-rate ila. et år. Derfor hentes ingen andre nøkler.

```
1 df = spark.sql("""
2     SELECT
3         ksa.PERNR_SUR AS PERNR
4         ,TO_DATE(dat.BEGDA, 'yyyyMMdd') AS DATE
5         ,dat.STAT2
6         ,dat.MASSN
7         ,dat.MASSG
8         ,org.KOSTL
9         ,org.BTRTL
10        ,org.PERSG
11        ,org.STELL
12        ,CONCAT(dat.BEGDA, CAST(org.ORGHEH AS STRING)) ORGEH -- Surrugatkøkkel som peker på riktig versjon av organisasjonshierarkiet.
13        ,org.PERSK
14        ,org.PLANS
15        ,ans.GESCH
16        ,CAST(FLOOR(datediff(
17            TO_DATE(dat.BEGDA, 'yyyyMMdd'),
18            TO_DATE(CAST(UNIX_TIMESTAMP(CAST(ans.GBDAT AS VARCHAR(8)), 'yyyyMMdd') AS TIMESTAMP
19        )) / 365.2524 AS INT) AS AGE -- Alder ved aktuell dato
20        ,sal.BSGRD / 100 AS EMPCT -- Omregner stillingsprosent til desimaltall
21        ,IF(sal.BSGRD < 100, 0, 1) AS BSGRD -- Sjekker om stillingsprosent er mindre enn 100. Hvis ja: Deltid (0). Hvis nei: Heltid(1).
22 FROM raw_pa0000 dat -- Personalendringer krever ikke duplisering i kildetabellen ettersom handlingen skjer på et spesifikt tidspunkt
23 JOIN keystore_ansatt ksa
24     ON dat.PERNR = ksa.PERNR_ORG
25 JOIN stg_pa0001 org
26     ON TO_DATE(dat.BEGDA, 'yyyyMMdd') = org.Date
27     AND dat.PERNR = org.PERNR
28 JOIN stg_pa0002 ans
29     ON TO_DATE(dat.BEGDA, 'yyyyMMdd') = ans.Date
30     AND dat.PERNR = ans.PERNR
31 LEFT JOIN stg_pa0008 sal
32     ON TO_DATE(dat.BEGDA, 'yyyyMMdd') = sal.Date
33     AND dat.PERNR = sal.PERNR
34 """)
35
36 df.write.format("delta").mode("overwrite").option("overwriteSchema", "True").saveAsTable("dfo_hr_lh.dwh_pa0000")
```

[2] ✓ -Command executed in 34 sek 488 ms by Hagen, Fredrik on 12:00:36 PM, 11/10/23

Fabric Lakehouse - dfo_hr_lh

Utforsker

dfo_hr_lh

Tables

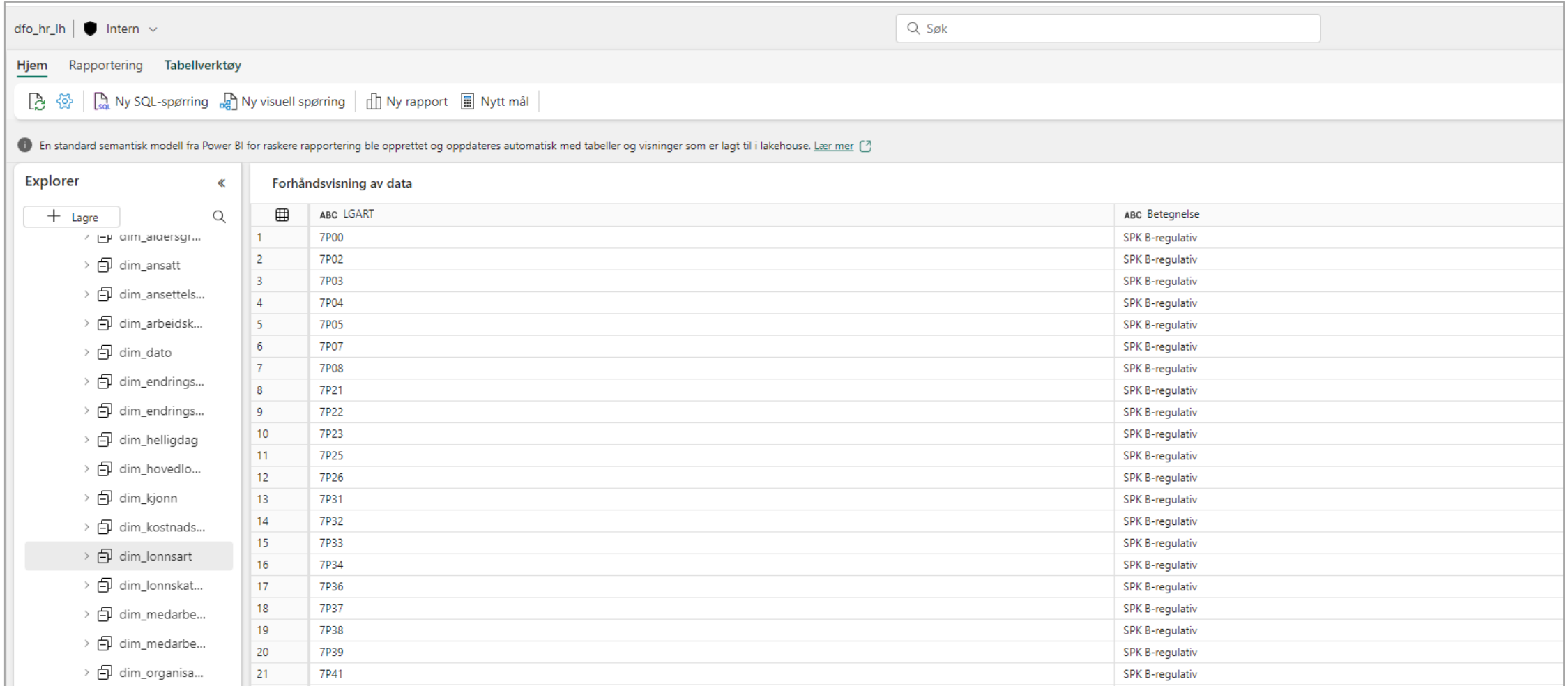
- date
- dwh_pa0000
- dwh_pa0002
- dwh_pa0007
- dwh_pa0008
- dwh_pa2001
- dwh_pa2002
- dwh_pa2010
- keystore_ansatt
- keystore_organisasio...
- raw_aldersgrupper
- raw_csks
- raw_holidays
- raw_hovedlonnstabell
- raw_hrp1000
- raw_hrp1001
- raw_hrp1051
- raw_pa0000
- raw_pa0001
- raw_pa0002

date

Viser 1000 rader

	Date	NumericDate
1	1/1/2009 12:00:00 AM	20090101
2	1/2/2009 12:00:00 AM	20090102
3	1/3/2009 12:00:00 AM	20090103
4	1/4/2009 12:00:00 AM	20090104
5	1/5/2009 12:00:00 AM	20090105
6	1/6/2009 12:00:00 AM	20090106
7	1/7/2009 12:00:00 AM	20090107
8	1/8/2009 12:00:00 AM	20090108
9	1/9/2009 12:00:00 AM	20090109
10	1/10/2009 12:00:00 AM	20090110
11	1/11/2009 12:00:00 AM	20090111
12	1/12/2009 12:00:00 AM	20090112
13	1/13/2009 12:00:00 AM	20090113
14	1/14/2009 12:00:00 AM	20090114
15	1/15/2009 12:00:00 AM	20090115
16	1/16/2009 12:00:00 AM	20090116
17	1/17/2009 12:00:00 AM	20090117
18	1/18/2009 12:00:00 AM	20090118
19	1/19/2009 12:00:00 AM	20090119
20	1/20/2009 12:00:00 AM	20090120
21	1/21/2009 12:00:00 AM	20090121
22	1/22/2009 12:00:00 AM	20090122
23	1/23/2009 12:00:00 AM	20090123
24	1/24/2009 12:00:00 AM	20090124
25	1/25/2009 12:00:00 AM	20090125
26	1/26/2009 12:00:00 AM	20090126
27	1/27/2009 12:00:00 AM	20090127
28	1/28/2009 12:00:00 AM	20090128

Endepunkt for SQL - dfo_hr_lh



The screenshot shows a Power BI report interface for the dataset 'dfo_hr_lh'. The top navigation bar includes 'Hjem', 'Rapportering', and 'Tabellverktøy'. Below this, there are icons for 'Ny SQL-spørring', 'Ny visuell spørring', 'Ny rapport', and 'Nytt mål'. A notification banner states: 'En standard semantisk modell fra Power BI for raskere rapportering ble opprettet og oppdateres automatisk med tabeller og visninger som er lagt til i lakehouse. [Lær mer](#)'.

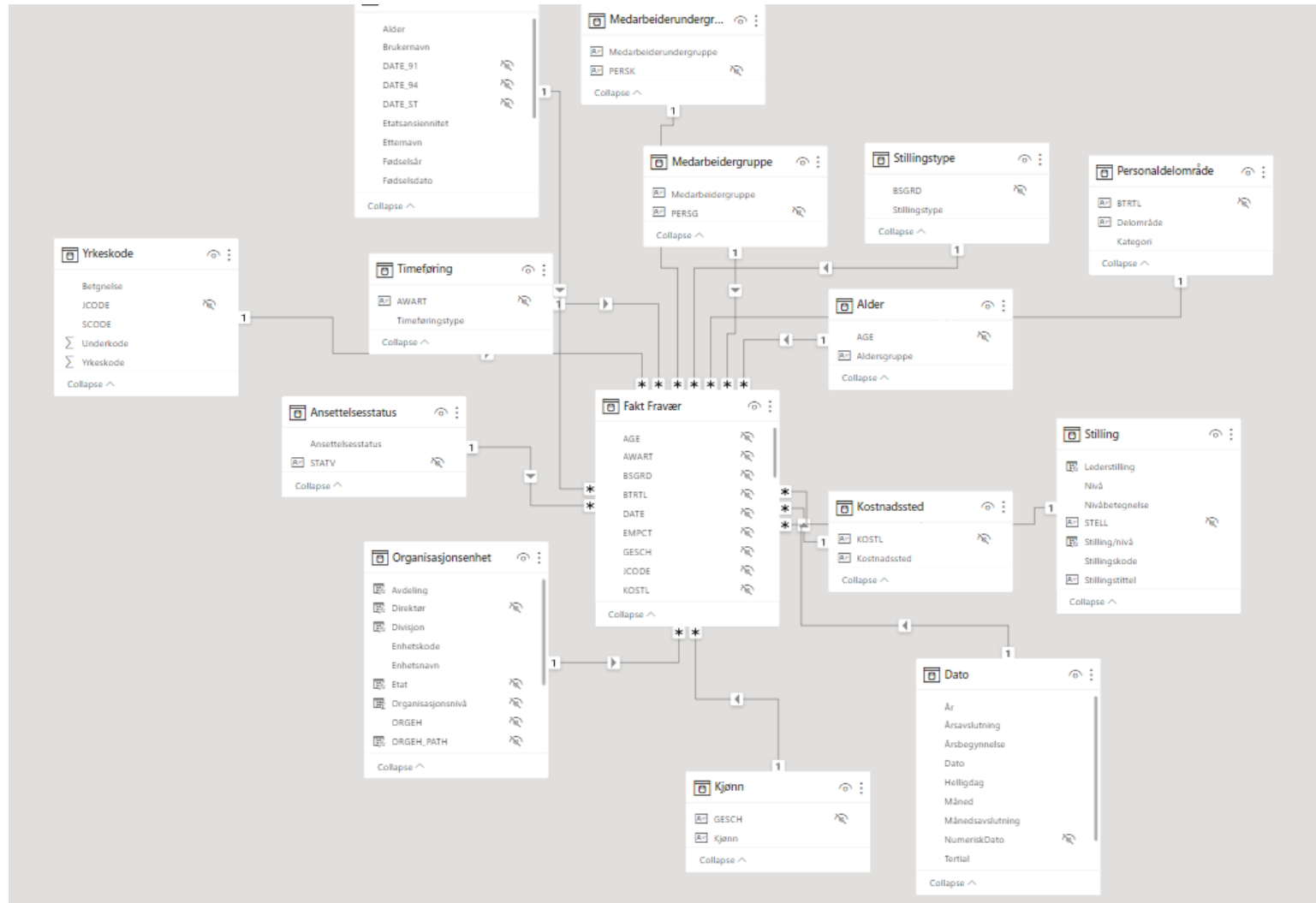
The 'Explorer' pane on the left lists various dimension tables, with 'dim_lønnsart' selected. The main area displays a 'Forhåndsvisning av data' table with the following columns: 'ABC LGART' and 'ABC Betegnelse'. The table contains 21 rows of data, each with a unique 'ABC LGART' value and the corresponding 'ABC Betegnelse'.

ABC LGART	ABC Betegnelse
1	SPK B-regulativ
2	SPK B-regulativ
3	SPK B-regulativ
4	SPK B-regulativ
5	SPK B-regulativ
6	SPK B-regulativ
7	SPK B-regulativ
8	SPK B-regulativ
9	SPK B-regulativ
10	SPK B-regulativ
11	SPK B-regulativ
12	SPK B-regulativ
13	SPK B-regulativ
14	SPK B-regulativ
15	SPK B-regulativ
16	SPK B-regulativ
17	SPK B-regulativ
18	SPK B-regulativ
19	SPK B-regulativ
20	SPK B-regulativ
21	SPK B-regulativ

Det er kun laget views på dimensjonstabellene av hensyn til ytelse.

Det er også kun dimensjonstabellene og measures sluttbrukerne skal forholde seg til i den semantiske modellen.

Eksempel på datamodell (1 av 6 faktabeller)





TOLLETATEN