

## Vedlegg 04 – Krav til ETL verktøy

Det er behov for et ETL-verktøy (Extract, Transform, Load) i løsningen med funksjonalitet som dekker hele livssyklusen for datahåndtering – fra innhenting til lagring og analyse.

### 1. Extract (Innhenting av data)

- **Støtte for flere datakilder:**
  - Trafikksensorer (f.eks. induktive sløyfer, radar, kamera).
  - GPS-data fra kjøretøy eller mobilapper.
  - API-er fra offentlige transportmyndigheter.
  - Historiske databaser og CSV/JSON-filer.
- **Streaming og batch-prosessering:**
  - Håndtering av sanntidsdata (Kafka, MQTT).
  - Periodisk innhenting av historiske data.

### 2. Transform (Databehandling og kvalitetssikring)

- **Datavask og validering:**
  - Fjerne duplikater, korrigere feil, håndtere manglende verdier.
- **Normalisering og standardisering:**
  - Konvertering til felles format (f.eks. tidssoner, enheter).
- **Aggregering og berikelse:**
  - Beregning av gjennomsnittshastighet, trafikkvolum, reisetid.
  - Kunne aggregere basert på ikke-sammenhengende tidsdata/attributter. F.eks. "Alle mandager i oktober per time"
  - Geokoding (knytte data til kartkoordinater).
- **Business rules:**
  - Klassifisering av trafikkflyt (f.eks. grønn/gul/rød), kjøretøy (tung/lett, lang/kort, el/bensin)
  - Identifisering av hendelser (ulykker, kø, påskeuke)

### 3. Load (Lagring og tilgjengeliggjøring)

- **API-eksponering:**
  - REST/GraphQL for tredjepartsbruk.

## 4. Overvåkning og administrasjon

- **Jobbplanlegging og orkestrering:**
  - Automatiserte pipelines
- **Feilhåndtering og logging:**
  - Varsling ved databrudd eller systemfeil.
- **Skalerbarhet og ytelse:**
  - Optimalisering for store datamengder.

## 5. Spesifikke funksjoner for trafikkdata

- **Sanntidsanalyse:**
  - Prediksjon av trafikkflyt og reisetid.
- **Historisk trendanalyse:**
  - Identifisering av mønstre (rushtrafikk, sesongvariasjoner).
- **Integrasjon med karttjenester:**
  - Visualisering på GIS-plattformer.