Samenwerkingsverband
van decentrale
OV-autoriteiten

Abstract van architectuur

Dataplatform OV



1. Abstract architectuur dataplatform OV

1.1. Inleiding

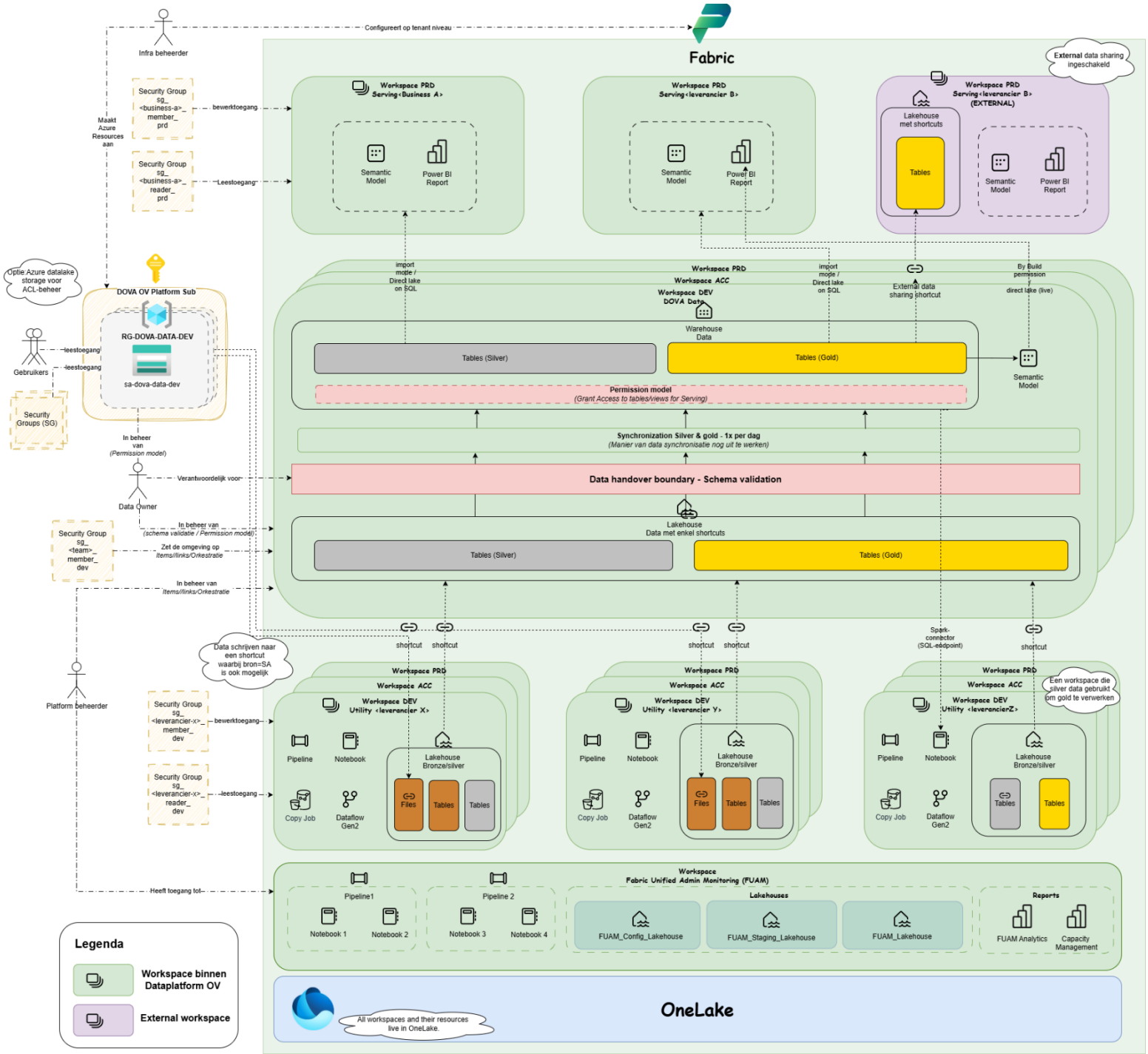
Het dataplatform wordt gerealiseerd in Microsoft Fabric en heeft als doel data uit meerdere bronnen en van verschillende leveranciers centraal samen te brengen. Hierdoor wordt het mogelijk om:

- data verder te verwerken en te verrijken;
- de kwaliteit van de data te borgen;
- data op een veilige en gecontroleerde manier te ontsluiten;
- het platform effectief te beheren en te onderhouden.

Daarnaast stelt het dataplatform gebruikers in staat om de verrijkte data te benutten voor analyses, rapportages en onderbouwde besluitvorming.

1.2. Algemeen overzicht

Het ontwerp van het dataplatform in Fabric wordt hieronder weergegeven. De uitwerking van de workspacestructuur (Utility/Data/Serving), het pull-synchronisatiemodel, autorisatie en de centrale opslag/publicatielaag volgen in het vervolg van de aanbesteding.



1.3. Medallion-architectuur

Het dataplatform is ingericht volgens de medallion-architectuur. Binnen deze architectuur wordt het ELT-principe (Extract, Load, Transform) toegepast, waarbij data eerst onbewerkt wordt geladen en vervolgens stapsgewijs binnen het platform wordt getransformeerd.

In tegenstelling tot ETL (Extract, Transform, Load) worden transformaties niet vóór, maar binnen het platform uitgevoerd. De data worden opgeslagen in verschillende logische lagen (Bronze, Silver en Gold) binnen de OneLake-opslag van Microsoft Fabric. Deze gelaagde opzet ondersteunt hergebruik, schaalbaarheid, kwaliteitsborging en governance.

- **Bronze**
  Bevat ruwe data die rechtstreeks wordt ingelezen vanuit externe en interne databronnen. In deze laag vindt nog geen of minimale bewerking plaats.

- **Silver**
  Data uit de Bronze-laag wordt opgeschoond, gevalideerd en genormaliseerd. Deze laag bevat betrouwbare en gestandaardiseerde data die geschikt is voor verdere verwerking.

- **Gold**
  Data uit de Silver-laag wordt verder verrijkt, geaggregeerd en aangevuld met businesslogica. Deze laag vormt de basis voor analyse, rapportages en besluitvorming.

1.4. Multi-vendorprincipe

Utility
Het dataplatform is zodanig ontworpen dat meerdere leveranciers geïsoleerd kunnen aansluiten om hun eigen dataverwerking uit te voeren binnen een separate Utility-workspace. Een leverancier kan binnen een afgebakende omgeving data ophalen vanuit externe en interne databronnen en deze verwerken binnen zijn eigen omgeving.

Hierbij is de centrale opslag van alle data expliciet gescheiden van de Utility-workspaces. Leveranciers zijn binnen hun Utility-workspace zelfstandig verantwoordelijk voor de verwerking van data. Deze verwerking strekt zich uit van het inlezen van brondata in de Bronze-laag, het opschonen en valideren van data in de Silver-laag tot het toevoegen van businesslogica in de Gold-laag, zonder (directe) interferentie met andere leveranciers.

Door het hanteren van uniforme datamodellen, naamgevingsconventies en vastgestelde kwaliteitsafspraken, die als richtlijnen aan leveranciers worden verstrekt, wordt geborgd dat de door leveranciers verwerkte data geschikt is voor centrale afname en verdere ontsluiting binnen het dataplatform.

Serving
Het dataplatform stelt daarnaast andere gebruikers in staat om gebruik te maken van specifieke datatabellen voor analyse- en rapportagedoeleinden in een eigen Serving-workspace. Ook deze gebruikers werken binnen geïsoleerde omgevingen, zodat gebruikers en gebruikersgroepen zelfstandig rapportages kunnen ontwikkelen op basis van een afgebakend en toegewezen deel van de data.

Binnen deze omgevingen hebben gebruikers bovendien de mogelijkheid om zelf datamodellen te ontwerpen en aan te passen, zonder invloed op andere gebruikers of datasets.

1.5. Gecentraliseerde dataomgeving

De centrale Data-workspace fungeert als de Single Source of Truth (SSOT) voor het volledige dataplatform en vormt de primaire publicatie- en ontsluitingslaag voor data. Vanuit deze omgeving krijgen eindgebruikers en leveranciers gecontroleerde toegang tot de beschikbare datatabellen.

Eindgebruikers die data nodig hebben voor analyse- of rapportagedoeleinden hoeven hiervoor niet rechtstreeks bij één of meerdere leveranciers aan te kloppen. Alle benodigde data zijn centraal beschikbaar gesteld via het dataplatform, wat zorgt voor eenduidigheid, herbruikbaarheid en consistente interpretatie van data.

1.6. Design principes

Binnen Microsoft Fabric bestaan beperkingen waardoor extra handelingen nodig zijn om het multi-vendorprincipe in combinatie met een centrale Data-workspace te realiseren. Daarnaast worden specifieke eisen en wensen gesteld aan het dataplatform, waaraan binnen Microsoft Fabric moet worden voldaan. De architectuur van het platform volgt uit de vastgestelde designprincipes.

Scheiding van verantwoordelijkheden
Taken en verantwoordelijkheden worden zowel logisch als technisch gescheiden, zoals ingestie, transformatie, opslag, semantiek en consumptie.
Dit draagt bij aan betere onderhoudbaarheid, schaalbaarheid en het voorkomen van ongewenste afhankelijkheden.

Isolatie van externe partijen
Externe partijen werken uitsluitend binnen strikt gescheiden, projectspecifieke workspaces. Hiermee wordt de integriteit, veiligheid en stabiliteit van de centrale Data-workspace gewaarborgd.

Single Source of Truth (SSOT)
Voor elke entiteit, definitie en KPI bestaat één centrale en gevalideerde waarheid.
Afgeleide datasets en rapportages verwijzen altijd naar deze bron en bevatten geen eigen interpretaties.

Standaardisatie
Datamodellen, naamgevingsconventies, KPI-definities, datatypes en ontwikkelpatronen worden gestandaardiseerd. Dit bevordert herbruikbaarheid, versnelt ontwikkeling en maakt oplossingen begrijpelijk en overdraagbaar.

Security by Design
Beveiliging is een integraal onderdeel van het ontwerp en wordt zoveel mogelijk op één centrale plek ingericht. Toegangscontrole, autorisaties (bijv. RLS/OLS) en auditing worden uniform toegepast volgens het least-privilege-principe.

Virtualisatie waar mogelijk
Om onnodige dataverspreiding tegen te gaan wordt data, waar mogelijk, aangeboden via gevirtualiseerde lagen, zoals views of shortcuts. Fysieke duplicatie van data wordt alleen toegepast wanneer dit aantoonbare voordelen biedt op het gebied van performance of stabiliteit.

Datakwaliteit als integraal onderdeel
Datakwaliteit is geen eenmalige activiteit, maar een vast onderdeel van het ontwerp van elke datapijplijn. Door middel van geautomatiseerde quality-audits worden validaties, controles en monitoring structureel ingericht en inzichtelijk gemaakt.

Loose Coupling
Componenten en workspaces binnen het dataplatform zijn zo los mogelijk gekoppeld. Wijzigingen in één onderdeel hebben daardoor minimale impact op andere onderdelen, wat flexibiliteit, onderhoudbaarheid en onafhankelijk ontwikkelen bevordert.

Platform boven project
Het dataplatform wordt ontworpen als een generiek en herbruikbaar platform, niet als een verzameling projectspecifieke oplossingen. Projectspecificiteit wordt zoveel mogelijk beperkt tot configuratie en data-inrichting.

Configuratie boven maatwerk
Waar mogelijk worden functionaliteiten gerealiseerd via configuratie, standaardpatronen en Infrastructure as Code (IaC), in plaats van maatwerkcode. Dit verhoogt onderhoudbaarheid, overdraagbaarheid en leidt tot lagere beheerkosten.

Traceability & Lineage
Data is te allen tijde herleidbaar naar de bron, het verwerkingsproces en de toegepaste transformaties. Lineage-informatie en metadata maken inzichtelijk waar data vandaan komt, hoe deze is gewijzigd en waar deze wordt gebruikt.

Productiestabiliteit boven previewfunctionaliteit
Het dataplatform maakt uitsluitend gebruik van algemeen beschikbare en productierijpe (Generally Available) functionaliteit. Public preview- of experimentele features worden niet ingezet in kritieke dataprocessen vanwege risico's op instabiliteit, functionele wijzigingen en beperkte ondersteuning.
Nieuwe platformfunctionaliteit wordt pas toegepast nadat deze productierijp is verklaard en aantoonbaar voldoet aan de eisen voor stabiliteit, beveiliging en beheer.

Evolutionair ontwerp
Het dataplatform is zodanig ontworpen dat nieuwe eisen, technologieën en functionaliteiten geleidelijk kunnen worden toegevoegd zonder grootschalige herstructurering of verstoring van bestaande functionaliteit.