



Bijlage 09 Programma van Eisen	
Status	Definitief 28 maart 2024

Inhoud

1	Inleiding	4
1.1	Doel.....	4
1.2	Doelstelling van het datawarehouse (DWH).....	4
1.3	Beschrijving huidige oplossing	4
1.4	Kenmerken en statistieken huidige oplossing	7
1.4.1	<i>ETL/ELT-proces</i>	7
1.4.2	<i>Staging area</i>	8
1.4.3	<i>Modelled store (Datavault)</i>	10
1.4.4	<i>Business store (Datamart)</i>	10
1.4.5	<i>Rapportages</i>	11
1.4.6	<i>Analyses en voorspelmodellen</i>	12
1.5	Datamanagement, data governance en datakwaliteit	15
1.6	Anonimisering persoonsgegevens	15
1.7	Doorontwikkeling datawarehouse.....	16
1.8	Doel programma van eisen	17
2	Algemene vereisten	18
2.1	Governance	18
2.2	Vergoeding (commerciële eisen)	20
2.3	Service Level Agreement.....	21
2.4	Beheer, ondersteuning en wijzigingen.....	22
2.4.1	<i>Beschikbaarheid</i>	22
2.4.2	<i>Herstelbaarheid</i>	23
2.4.3	<i>Incidentbeheer</i>	24
2.4.4	<i>Wijzigingsbeheer en Onderhoud</i>	25
2.5	Prestaties en schaalbaarheid	25
2.6	Informatiebeveiliging en privacy	27
3	Functionele vereisten	32
3.1	Gegevensintegratie.....	32
3.2	Gegevensanalyse en rapportage	34
3.3	Gegevensbeheer	36
3.4	Ongestructureerde data.....	37

4	Technische vereisten	38
5	Eisen ten aanzien van continuering	40
6	Doorontwikkeling datawarehouse	42
7	Exit criteria	45

1 Inleiding

1.1 Doel

Het doel van dit document is het opstellen van een programma van eisen voor de continuering van de Datawarehousedienstverlening van het Vervangingsfonds/Participatiefonds (hierna: VfPf). Deze eisen zijn bedoeld om de functionele en technische vereisten van het systeem te definiëren en ervoor te zorgen dat de SaaS-oplossing en de daaraan gerelateerde inzet van personen voldoet aan de behoeften van de organisatie.

1.2 Doelstelling van het datawarehouse (DWH)

DWH is een ondersteunde applicatie voor de reguliere bedrijfsprocessen van VfPf om te zorgen dat de dienstverlening aan schoolbesturen wordt verbeterd. Data wordt samengebracht om - in de vorm van rapportages - betrouwbare informatie (o.a. trendanalyses en forecasts) te genereren zodat de doelgroep tijdig en beter kan anticiperen op toekomstige ontwikkelingen en behoeften.

Twee belangrijke doelstellingen van VfPf zijn:

- We gebruiken data continu in onze dienstverlening en om te verbeteren en te ontwikkelen;
- Vanuit onze unieke positie zijn wij in staat om relevante kennis, informatie, onderzoeken en data bij elkaar te brengen om zo de klant, de sector en onze eigen organisatie verder te helpen.

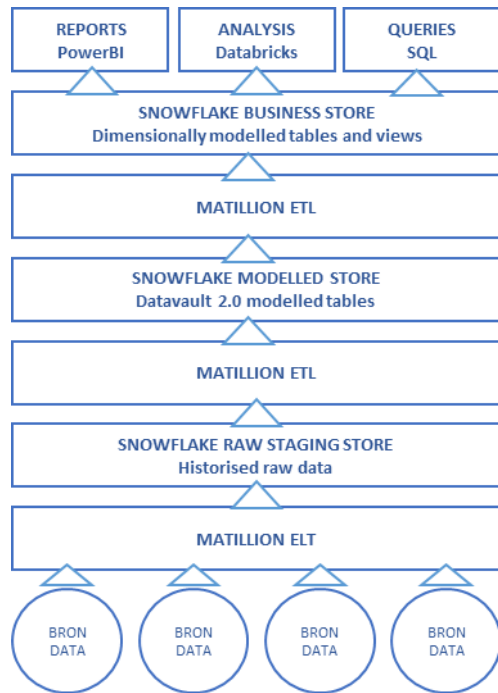
1.3 Beschrijving huidige oplossing

In 2019 is gestart met het ontwikkelen van een datawarehouse en business-intelligence-omgeving. Deze dienst is geheel opgezet als 'Software as a Service' (hierna: SaaS). De eerste versie van de dienst is in 2020 in productie genomen. Daarna hebben er continu ontwikkelingen plaatsgevonden in termen van:

- Aansluiten van nieuwe databronnen;
- Verfijning van de datamodellen;
- Nieuwe en aangepaste rapportages, analyses en voorspelmodellen;
- Aansluiten van Databricks (de huidige data-science-applicatie).

De huidige opzet van het datawarehouse is geheel gebaseerd op de Cloudproducten: Snowflake (Business Critical editie 8.2.0) voor de database (pre-commit 12k credits/jr), Matillion (voor ETL/ELT), PowerBI (8 licenties) voor rapportage en Databricks (Data Analysis) voor analyse en voorspelmodellen. De

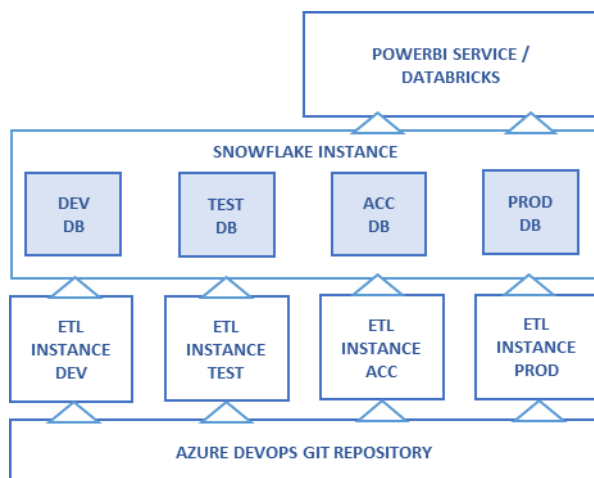
databasemodellering is volgens de principes van Datavault 2.0. De huidige oplossing is hieronder schematisch weergegeven.



De huidige technische oplossing bestaat uit vier omgevingen:

- Ontwikkeling;
- Test;
- Acceptatie;
- Productie.

De rapportage-omgeving inclusief Databricks is alleen aangesloten op de acceptatie- en productieomgeving. VFPf heeft zelf geen toegang tot de ontwikkel- en testomgeving.



Data load

De kern van dit onderdeel bestaat uit vier virtual machines in Microsoft Azure. Deze machines zijn onafhankelijk van elkaar en draaien hun eigen instance van Matillion ETL/ELT-tooling, een eigen connectie met databronnen, een eigen softwareversie en een eigen datadoel.

DWH

Er is één instance van Snowflake waar de vier omgevingen van de ontwikkelstraat worden gehost. Elke database is opgebouwd binnen het Snowflake-account en deze zijn strikt van elkaar gescheiden. Een account dat toegang heeft tot een ontwikkeldatabase kan bijvoorbeeld nooit toegang hebben tot een acceptatie- of productiedatabase.

Momenteel wordt alleen Snowflake gebruikt als databasemanagementsysteem (DBMS) voor het datawarehouse. We maken geen gebruik van 'Snowflake shares'.

De Snowflake-rechten worden uitgedeeld op verzoek van Opdrachtnemer. Het beheer daarover gebeurt door (de huidige) leverancier. Er zijn custom roles binnen Snowflake uitgegeven.

Reporting (PowerBI)

PowerBI heeft een aparte test/acceptatieomgeving waarin nieuwe rapportages worden ontwikkeld. Na acceptatie worden rapportages naar de productieomgeving overgebracht. De regie van de deployments ligt in de meeste gevallen bij aangewezen domeineigenaren of datastewards binnen VfPf.

Dit betekent dat wanneer iets misgaat in verversing van PowerBI-rapportages, dit wordt gesignaleerd en opgelost door het datateam van VfPf. Mocht de oorzaak hiervan bij de leverancier van de SaaS-service liggen, dan wordt hiermee geschakeld.

PowerBI maakt gebruik van een 'imported dataset'. Omdat het datawarehouse zelf eens per nacht wordt bijgewerkt, is normaal geen behoefte om de gegevens in PowerBI meer dan één keer per dag te verversen.

Code repository

GitHub is de repository waarin de code centraal wordt beheerd. Zowel de code voor ETL/ELT (nu Matillion),

als de database-creatie staan in de repository. Het is mogelijk om ook de rapportages onder te brengen in de repository, maar dit gebeurt momenteel niet. Wij gebruiken momenteel niet 'GitHub Actions'.

1.4 Kenmerken en statistieken huidige oplossing

1.4.1 ETL/ELT-proces

Het ETL/ELT-proces is hieronder schematisch weergegeven.



De data wordt via een ELT-proces (Extract-Load-Transform) in de staging-omgeving geladen. Daarna wordt de data via twee ETL (Extract-Transform-Load) doorgeladen naar de gehistoriseerde 'modelled store' (Datavault) en de business store.

Matillion wordt ingezet voor het ETL/ELT-proces en Snowflake is het databasesysteem. Beide systemen zijn Software-as-a-Service (SaaS).

Het laadproces wordt gemonitord in Matillion. Afwijkingen en fouten worden automatisch gesignaleerd en doorgestuurd naar de beheerorganisatie.

Voor een aantal databronnen is de periodieke dataversiersing geautomatiseerd. Voor een aantal wordt dit handmatig gestart als nieuwe gegevens beschikbaar zijn. Alle databronnen zijn 'pull'-bronnen. Het ETL/ELT-proces haalt dus zelf de data op. Er wordt geen data gepusht naar de staging area. De nachtrun wordt gestart om 1:00 uur en draait momenteel ongeveer 6,5 uur. We verwachten dat dit in de toekomst langer wordt door meer en groeiende databronnen.

Voor zover de brongegevens uit VfPf-systemen komen, zijn hier afspraken over gemaakt. We maken echter ook gebruik van openbare databronnen als DUO. Hier zijn over het algemeen geen afspraken over dataleveringen naar het datawarehouse gemaakt. Afspraken over dataleveringen naar de operationele systemen van VfPf vallen buiten deze context.

Voor de staginglaag van het datawarehouse wordt een generiek component gebruikt welke de ELT van de bron naar de staging-tabellen genereert, zowel voor de DDL als de load (DML). Voor de Datavault gebruikt de huidige opdrachtnemer standaarden en template-mappings als basis voor uitbreidingen.

1.4.2 Staging area

De 'staging bestaat' uit twee onderdelen:

- Pre-staging. Dit zijn de volledige gegevens van alleen de laatste upload van de databron.
- Staging. Hier wordt de laatste upload vergeleken met de vorige upload(s). Wanneer wijzigingen hebben plaatsgevonden wordt het oude record voorzien van een einddatum en een nieuwe aangemaakt. Op basis hiervan kunnen we altijd terughalen wat de vorige situatie was.

Onderstaande tabel bevat de relevante gegevens over de koppelingen met de bronsystemen.

STAGING

Bron	Bevat voornamelijk	Koppeling	Indicatie aantal tabellen	Grootte in Gb	Full load / Delta	Laadfrequentie	Opmerking
Parflex (Microsoft Dynamics)	Klanten (Participatiefonds), uitkeringsgerechtigden, uitkeringen, betalingen van uitkeringen, re-integratiedossiers en -activiteiten.	SQL Server (ODS)	40		Full load	Dagelijks	Wordt 2024 uitgefaseerd, maar data moet in DWH blijven.
CRM (Microsoft Dynamics)	Klanten (Participatiefonds), onderlinge relaties en contactmomenten.	SQL Server (ODS)	20		Full load	Dagelijks	
Fenix	Uitkeringen, betalingen op uitkeringen en facturen naar schoolbesturen op basis van deze betalingen.	API (Json)	20		Full load	Dagelijks	
MijnVF	Klanten (Vervangingsfonds), aanstellingen, verzuimgegevens, vervangingsdeclaraties, vervangingsproducten, vervangingspoolen.	XML-bestanden via SFTP	100		Full load	Maandelijks	
Perspectief	Klanten (Participatiefonds), uitkeringsgerechtigden, re-integratiedossiers en -activiteiten en statusovergangen binnen de re-integratieprocessen.	API (Json)	4		Full load	Dagelijks	Tijdelijke koppeling met minimale tabellen. Uiteindelijke koppeling wordt 2024 Q1/Q2 gelegd
Arbomeester	Respons, vragen en antwoorden vanuit periodieke risicoinventarisatie en -evaluatie (RI&E).	SQL Server (ODS)	25		Gedeeltelijk full, gedeeltelijk delta	Dagelijks	
DUO basisgegevens scholen en besturen	Stamgegevens scholen en schoolbesturen, onderlinge relaties en overgangen.	API (Json)	3		Full load	Wekelijks	Automatisch ophalen van website
DUO Caribisch	Stamgegevens scholen en schoolbesturen in het Caribisch Gebied	API (Json)	1		Full load	Wekelijks	
DUO Financiën	Detailgegevens balans en V&W van schoolbesturen	CSV via website	100		Full load	Jaarlijks	Handmatig ophalen van website
DUO Leerlingen	Leerlingstatistieken per school per jaar	CSV via website	25		Full load	Jaarlijks	Handmatig ophalen van website
DUO Onderwijsinspectie	Oordelen en toezichtarrangementen van scholen	CSV via website	3		Full load	Jaarlijks	Handmatig ophalen van website
Klanttevredenheids onderzoeken (KTO)	Respondenten, enquêtevragen en -antwoorden	API (Json)	10		Full load	Dagelijks	
Referentie- en mastertabellen	Referentiegegevens en groeperingen specifiek voor het DWH	Excel	20		Full load	Bij wijziging	Ongeveer 5 keer per jaar
CBS schoolwegingen	Schoolwegingen per school en per vestiging	Excel via website	1		Full load	Jaarlijks	Handmatig ophalen van website
Twinfield (boekhoudsysteem)	Interne financiële gegevens	API (Json)	55		Full load van alleen open jaar	Dagelijks	Afgesloten boekjaren worden niet meer dagelijks ingelezen

Een indicatie van het aantal rijen dat in de 'Staging' worden ingelezen staat hieronder.

Dagelijks: 23.000.000 rijen
Additioneel wekelijks: 250.000 rijen
Additioneel maandelijks: 3.700.000 rijen
Additioneel jaarlijks: 950.000 rijen

Deze rijen worden daarna verwerkt in de daaropvolgende stadia.

1.4.3 Modelled store (Datavault)

Het huidige Datavault-model is grotendeels gebaseerd op de bronsystemen. Er zijn wel basisobjecten (Hubs) welke vanuit meerdere bronnen worden gevuld. Daar waar objecten uit meerdere bronsystemen dezelfde functionele betekenis hebben, komt dit ook als één object in de Datavault terug.

Er is nog geen formeel business objectentabel/enterprise datamodel binnen VfPf. Deze is wel in ontwikkeling.

De modelled store (Datavault 2.0) bevat momenteel ongeveer 350 tabellen. Hier zijn de tabellen uit twee in ontwikkeling zijnde nieuwe applicaties Perspectief en Arbomeester nog niet volledig in meegenomen:

MODELLED STORE

Soort	Aantal tabellen	Aantal records	Grootte in Gb
Hubs	56	66.000.000	22
Links	80	73.000.000	70
Satellites	212	403.000.000	113
Overig	18	480.000	< 1

De tabelstructuur van de datamarts (hieronder beschreven) ontstaat uit de datavaulttabellen.

1.4.4 Business store (Datamart)

In de business store zijn de gegevens dimensioneel gemodelleerd in een sterschema, dus met feiten en dimensies. We gebruiken alleen bij hoge uitzondering sneeuwvlokdimensies.

De business store betreft de volgende businessdomeinen:

- Schoolbesturen, scholen en vestigingen;
- Aanstellingen, verzuim en vervanging;
- Re-integratie;
- Uitkeringen en betalingen;
- Klantcontacten;

- Financiële transacties;
- Klanttevredenheidsonderzoeken (vragenlijsten).

Datarelevante statistieken zijn in onderstaande tabel opgenomen. Hier zijn de wijzigingen en toevoegingen vanuit Perspectief en Arbomeester, twee in ontwikkeling zijnde nieuwe applicaties nog niet volledig in meegenomen.

BUSINESS STORE

Soort	Aantal tabellen	Aantal records	Grootte in Gb
Feiten	28	42.000.000	32
Dimensies	55	9.000.000	5
Rapportagetabellen	3	55.000.000	57

De huidige Opdrachtnemer biedt voldoende documentatie en ondersteuning voor het Data Team om te weten welke tabelkoppelingen men moet maken om het juiste resultaat te krijgen. Er is zeer gedetailleerde documentatie beschikbaar in standaardformaten als Word, Excel en Powerpoint. Deze wordt momenteel gedeeld via Teams. Er is geen historisering van de documentatie.

Het gebruik van het datawarehouse vergt daarnaast kennis van SQL en enige kennis van datamodellering (sterschema). Bij het Data Team is het kennisniveau hiervoor op niveau. Er zijn hierin momenteel geen opleidingen of trainingen gewenst.

Iedereen binnen het Data Team maakt gebruik van dezelfde sterschematabellen. Er is geen datamart per gebruiker of gebruikersgroep.

In eerste instantie is de modellering van de datamarts voornamelijk op basis van de expertise van de huidige Opdrachtnemer gebeurd. Momenteel krijgen we als VfPf steeds meer zelf kennis hiervan en trekken we hierdoor meer modellering naar ons toe.

1.4.5 Rapportages

Rapportages worden door het Data Team binnen VfPf ontwikkeld. Dit gebeurt voornamelijk met behulp van PowerBI, zeker als het gaat om periodieke standaardrapportages. Op dit moment zijn dit er zo'n 30. Standaardrapportages zijn gebaseerd op de gegevens in de business store. Deze zijn voornamelijk gericht op de volgende onderwerpen:

- Aanmelding- en presentielijsten van bijeenkomsten;
- Basisinzichten schoolbesturen en onderwijsinstellingen;
- Kosten versus begroting van VfPf-diensten en -producten;
- Redenen instroom WW (beëindigingsgronden);

- Financiële rapportages;
- Berekeningstool verwachte uitkeringskosten voor schoolbesturen;
- Lijsten voor uitnodigingen voor klanttevredenheidsonderzoeken;
- KPI-rapportages voor uitvoerders van VfPf;
- Prestatie-indicatoren van het Vervangingsfonds (voor OCW en de sociale partners);
- Uitkeringen per schoolbestuur (voor de regiocoördinatoren VfPf);
- Resultaten re-integratieactiviteiten;
- Analyse 'Stille reserves' (bevoegde leerkrachten die niet actief in het onderwijs zijn);
- Basisgegevens uitkeringen primair onderwijs;
- Verzuimrapportage;
- Rapportage vervangingen;
- Klantcontacten;
- Managementdashboards.

Voor ad-hoc-rapportages wordt ook regelmatig de Snowflake-SQL-client in combinatie met Excel ingezet. Hier kan ook de staging store of de modelled store als gegevensbron dienen. Vaak wordt deze oplossing gekozen bij onderzoeken naar datakwaliteit en eenmalige overzichten van gegevens die (nog) niet zijn gemodelleerd in de business store. Het Data Team heeft dus toegang tot alle drie de data stores.

Semantische modellen binnen PowerBI worden door het Data Team zelf gemaakt op basis van de informatiebehoefte van het PowerBI-rapport. Beheer en onderhoud hiervan gebeurt ook door het Data Team zelf.

Een deel van de rapportages wordt dagelijks ververs. Het piekmoment van de verversing van PowerBI-rapportages ligt momenteel tussen 8:00 uur en 9:00 uur. Waarbij ongeveer de helft gestart wordt tussen 8:30 uur en 9:00 uur 's ochtends.

1.4.6 Analyses en voorspelmodellen

Analyses en voorspelmodellen worden uitgevoerd met behulp van Databricks, Excel en PowerBI.

Databricks

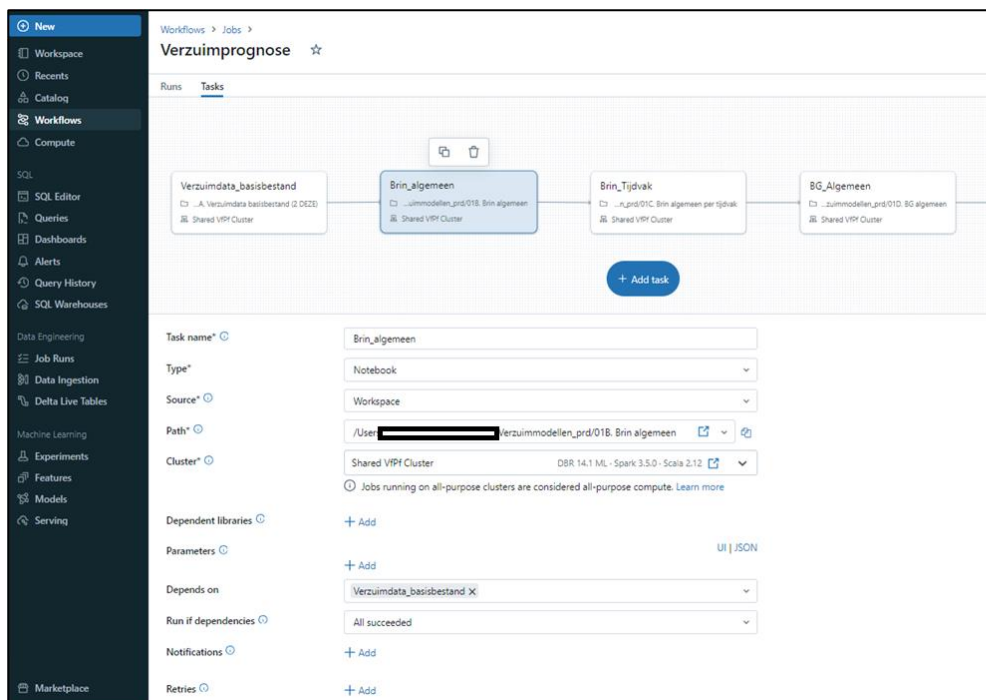
In Databricks worden SQL-query's rechtstreeks uitgevoerd op het datawarehouse. De data wordt vervolgens opgeslagen in Hadoop Distributed File System (HDFS), een gedistribueerd bestandssysteem voor opslag van grote hoeveelheden gegevens.

Verdere datapreparatie, analyses en voorspelmodellen worden uiteindelijk uitgevoerd met behulp van Notebooks die opgeslagen worden in de 'Workspace' in Databricks. Een Databricks Notebook is een interactief document dat wordt gebruikt om code te schrijven, uit te voeren en te delen. Het is een webgebaseerde interface waarin gebruikers code kunnen schrijven in verschillende programmeertalen, zoals Python, Scala, R en SQL en lijkt sterk op een Python Notebook. In de praktijk wordt momenteel met name in Python en SQL geprogrammeerd. Voor de toekomst willen we de mogelijkheid openhouden om ook in R te programmeren.

Een belangrijk hulpmiddel voor de ontwikkeling van voorspelmodellen vormt de functionaliteit 'AutoML' van Databricks. AutoML is een geautomatiseerd machine-learning-framework dat ontworpen is om het proces van modelontwikkeling te vereenvoudigen en te versnellen. Het stelt gebruikers in staat om geavanceerde machine-learningmodellen te bouwen en te implementeren zonder deze te hoeven programmeren. Deze functionaliteit wordt door de Data Scientists van VfPf gebruikt om snel en efficiënt initiële basismodellen te ontwikkelen die daarna gefinetuned worden. Deze manier van werken kost aanzienlijk minder tijd dan het vanaf de basis programmeren.

In Databricks kunnen voorspelmodellen automatisch geïmplementeerd en beheerd worden en dit is voor VfPf een belangrijke functionaliteit. Modellen kunnen eenvoudig op nieuwe data uitgescoord worden door middel van een automatisch gegenereerd notebook dat desgewenst handmatig aangepast kan worden, wat optimale flexibiliteit biedt.

Via Databricks 'Workflows' kunnen taken geautomatiseerd uitgevoerd worden. Hierin worden taken voor ETL-, analyse- en machine learning gedefinieerd, beheerd en gemonitord. Een voorbeeld van een Workflow staat hieronder afgebeeld. Zie voor meer informatie: <https://www.databricks.com/product/workflows>.



Voor meer informatie over Databricks verwijzen we naar <https://www.databricks.com/>. Hier is ruim voldoende documentatie te vinden en kan het platform gedurende een gratis proefperiode uitgetest worden.

Specifiek wordt nu het volgende afgenomen m.b.t. Databricks: Azure Databricks - Premium - All-purpose Compute. De onderdelen (1) Jobs Compute, (2) Delta Live Tables, (3) Databricks SQL en (4) Model Serving vallen allemaal onder de All-purpose Premium Compute variant. Aantal units in het afgelopen kwartaal (november t/m januari) was gemiddeld 1.264 per maand (met in november het grootste aantal van 2.234 units). Kosten worden via Azure marketplace afgenomen en afgerekend.

Momenteel zijn er twee 'All-purpose compute' clusters en een 'Serverless SQL Warehouse' die gebruikt worden:

- Cluster 1: workertype DS3V2 14 GB memory, 4 cores, 2-10 workers; driver type: DS13V2: 56 GB memory, 8 cores
- Cluster 2: DS3V2 14 GB memory, 4 cores
- SQL Warehouse: 2x-Small (4 DBU/h)

Excel

In de Snowflake-SQL-client wordt door middel van een SQL-query de benodigde data uit het datawarehouse gehaald. Vervolgens wordt de data geëxporteerd naar een CSV-bestand dat ingelezen wordt in Excel voor verdere analyse.

PowerBI

De tabellen uit het datawarehouse worden geïmporteerd in PowerBI en daar aan elkaar gekoppeld. Vanuit PowerBI kunnen gemaakte visualisaties en tabellen worden geëxporteerd.

1.5 Datamanagement, data governance en datakwaliteit

De huidige aanbieder heeft een uitgebreid dashboard waarin datakwaliteitsissues worden gesignaleerd. Deze signaleringen hebben voornamelijk betrekking op de datakwaliteitsdimensies Geldigheid en Unicité. Wanneer dit direct oplosbaar is door de aanbieder wordt dit ook door de aanbieder gedaan. Waar nodig wordt VfPf op de hoogte gehouden. Andere dimensies van datakwaliteit (Consistentie, Nauwkeurigheid, Volledigheid en Tijdigheid) worden over het algemeen opgemerkt door het Data Team en/of de eindgebruikers van de rapportages. Op deze datakwaliteitsdimensies willen we graag stappen maken. De datakwaliteit kunnen we op dit moment niet kwantificeren. Dit zal een van de eerste stappen zijn in een datakwaliteitstraject.

Waar mogelijk worden proceseigenaren verzocht de datakwaliteitsissues op te lossen in de betreffende bronsystemen. Procedures hierin zijn echter niet formeel vastgelegd. Verbetering van de datakwaliteitseisen, -signalering, en -oplossingsprocedures staan op de agenda van 2024.

Voor wat betreft data governance zijn er geen formele overleggen. Vanuit zowel ICT als het Data Team zijn initiatieven gestart om de datavolwassenheid te vergroten en data governance en datakwaliteit een prominentere plek te geven. Deze zijn op dit moment verkennend van aard. VfPf is zich ervan bewust dat dit een intensief en organisatiebreed veranderproces is.

De logging vanuit Azure geeft waarschuwingen omtrent security risks welke door de leverancier worden gesignaleerd en opgepakt. Op basis hiervan is en wordt het platform nog steeds veiliger gemaakt. Audit logging vindt alleen voor single-sign-on accounts plaats. Er bestaan gedeelde accounts (bijvoorbeeld Matillion). Hierdoor is logging naar individuele gebruikers in de huidige situatie nog niet altijd mogelijk.

Tot op zekere hoogte zijn functionele databeschrijvingen van het datawarehouse opgenomen in de gedetailleerde datamappings. Een bronsysteemoverstijgend enterprise datamodel inclusief datadictionary ontbreekt op dit moment. Het maken en onderhouden van in ieder geval een datadictionary is onderdeel van de plannen van 2024.

1.6 Anonimisering persoonsgegevens

In overleg tussen VfPf en de Opdrachtnemer is bepaald welke attributen moeten worden geanonimiseerd. Over het algemeen zijn dit attributen die direct of indirect herleidbaar zijn tot een individueel persoon. Anonimiseren gebeurt vervolgens in de ETL-omgeving op de Ontwikkeling-, Test- en Acceptatie-omgeving voordat de gegevens in de staging-laag komen. Hierdoor komen deze ook geanonimiseerd in de Datavault

en Datamart-laag. Op de productie-omgeving wordt geen anonimiseren of pseudonimisering toegepast. Dit proces wordt ook toegepast op nieuwe databronnen.

1.7 Doorontwikkeling datawarehouse

Het datawarehouse neemt een steeds belangrijkere plaats in binnen VfPf. Tegelijkertijd heeft VfPf de ambitie om steeds meer de rol van kenniscentrum binnen de sector Primair Onderwijs op zich te nemen. De ontwikkelingen op het datawarehouse staan daarom nooit stil:

- Nieuwe databronnen worden toegevoegd en bestaande databronnen worden vervangen;
- Nieuwe dataproducten en -diensten leiden tot uitbreidingen en verbeteringen van de business store (datamart);
- Datakwaliteit wordt steeds belangrijker en omvangrijker. De signaleringsfunctie van het datawarehouse richting de eigenaren van de bronsystemen neemt toe;
- Nieuwe rapportages en voorspelmodellen worden ontwikkeld. Bestaande worden verbeterd en uitgebreid. Er is nog geen concrete roadmap beschikbaar voor rapportages en voorspelmodellen. Er wordt momenteel gewerkt aan de implementatie van een verzuimprognosemodel op schoolniveau en daarnaast worden jaarlijks diverse voorspelmodellen (door)ontwikkeld voor de begroting van VfPf. Dit bestaat onder meer uit prognosemodellen voor verzuim- en vervangingskosten en prognosemodellen voor de werkeloosheidskosten en de grondslag in het primair onderwijs;
- Nieuwe manieren om data en rapportages te ontsluiten, bijvoorbeeld via een webportaal voor de klanten van VfPf;
- Meer gebruikers van het datawarehouse.

Het is van belang dat de overname van de Datawarehousedienst zo min mogelijk negatieve gevolgen heeft voor de continuïteit en ambities van VfPf.

Het Data Team bestaat uit acht personen: vier Data Analisten, een Business Analyst, een Data Specialist, en twee Data Scientists. Het team is tot op zekere hoogte zelfsturend en heeft mandaat voor ontwikkelingen die niet te ingrijpend of strategisch voor VfPf zijn.

Stakeholders voor wat betreft informatievoorziening zijn voornamelijk: de regiocoördinatoren, afdeling Control, de proceseigenaren en het managementteam maar ook OCW, het VfPf-bestuur en de sociale partners. Eisen en verzoeken rondom rapportages, analyses en voorspelmodellen worden gemanaged door de Business Analyst, de Data Analisten en de Data Scientists uit het Data Team. Stakeholders voor wat betreft de databronnen zijn ICT en de proceseigenaren. Eisen en verzoeken vanuit het DWH worden gemanaged door de Data Specialist uit het Data Team.

Het beheerteam en het ontwikkelteam van de Datawarehousedienst bestaan volledig uit medewerkers van de leverancier van deze dienst. Deze worden functioneel aangestuurd door de Data Specialist uit het Data Team. Het Data Team zal nauw samenwerken met het ontwikkelteam van de leverancier volgens een Agile-Scrum-aanpak, bij voorkeur onder aanvoering van een scrummaster van de leverancier. Bij projecten wordt het ontwikkelteam in principe via de Product Owner-rol (bij Opdrachtgever) betrokken maar dat kan ook rechtevreeks door medewerkers van de leverancier op te nemen in de Project Teams.

Het Data Team heeft een people manager en een inhoudelijk manager. Deze laatste neemt de beslissingen met grote en/of strategische impact en heeft hiervoor ook het mandaat.

1.8 Doel programma van eisen

Zoals hierboven is omschreven heeft VfPf twee doelen met de Datawarehousedienst:

1. Continuering van de dienst
VfPf wil het datawarehouse en de dienstverlening hier rondom continueren. Dat wil zeggen dat huidige rapportages, analyses, modellen en selecties periodiek verversi kunnen blijven worden en nieuwe rapportages gemaakt kunnen worden met betrouwbare en recente gegevens.
2. Doorontwikkeling van het datawarehouse
VfPf wil het datawarehouse blijven ontwikkelen door:
 - a. Meer interne en externe gegevens aan het datawarehouse toe te voegen;
 - b. De gegevensbetrouwbaarheid te vergroten;
 - c. De toepasbaarheid van de gegevens te vergroten.

Het programma van eisen dient als beschrijving voor de eisen die VfPf stelt aan de oplossing om beide doelen te bereiken, in termen van:

- Continuïteit & verbetering dienstverlening;
- Prestaties & Schaalbaarheid;
- Onderhoud & Beheer;
- Beveiliging en zorgvuldige, efficiënte en effectieve omgang met gegevens;
- Continuïteit ontwikkelteam in kennis en capaciteit.

2 Algemene vereisten

2.1 Governance

VfPf hanteert het volgende Governancemodel:



Het team van VfPf bestaat uit verschillende rollen welke acteren op strategisch, tactisch en operationeel vlak.

Het MT is verantwoordelijk voor het richtinggevend beleid en de strategie. De verantwoordelijkheid voor de implementatie daarvan ligt bij de procesverantwoordelijken en strategisch adviseur/architect. Voor de technische invulling van de datawarehousedienstverlening is de Proceseigenaar IT verantwoordelijk. De Key-user van de datawarehousedienstverlening is de Data Specialist en deze vervult daarmee de rol van de Product Owner (PO).

Op tactisch niveau vindt de aansturing plaats door de Product Owner en de Servicemanager. De Servicemanager is verantwoordelijk voor het contractmanagement. Beide rollen zijn nauw betrokken bij de ontwikkelingen en voeren het meest overleg met hun counterparts bij de Opdrachtnemer. De Servicemanager richt zich daarbij vooral op de contractuele en financiële afspraken, wederzijdse verplichtingen en op operationeel vlak de SLA en DAP. De PO richt zich op de functionele eisen en prestaties en stuurt het DWH-ontwikkelteam aan. Zowel de PO als de Servicemanager rapporteren aan de contracteigenaren binnen het MT en de Proceseigenaar IT.

Op operationeel vlak wordt het beheer door RUN gecoördineerd en wordt elke wijziging getest. Met RUN wordt het team bedoeld van Service delivery coördinatoren bij Opdrachtgever die de operationele processen aansturen, beheren en bewaken ofwel operationeel regie voeren.

Key-users en vertegenwoordigers namens onze klanten zullen desgewenst gericht worden toegevoegd aan onderdelen van het voortbrengingsproces. De gebruikers binnen VfPf zullen voornamelijk gebruikers zijn van inkoop- en servicemanagement.

- Eis 1: Communicatie en overleg tussen Opdrachtgever en Opdrachtnemer vindt plaats op diverse niveaus (operationeel, tactisch en strategisch). De Opdrachtnemer richt de overlegvormen in op operationeel, tactisch en strategisch niveau voor de beheerfase. In de ontwikkelfase wordt gebruik gemaakt van de overlegstructuren die passen binnen de werkwijzen van VfPf.
- Eis 2: VfPf stelt de volgende eisen aan de overlegstructuur:
- Ieder overleg vindt minimaal plaats in de bij de overlegvorm vast te stellen frequentie, waarbij de frequentie aangepast kan worden indien dit door Opdrachtgever noodzakelijk wordt geacht; Initieel vindt het operationeel (technisch) overleg iedere maand plaats, het tactisch overleg ieder kwartaal en het strategisch overleg ieder half jaar. Daarnaast vindt tweemaal per week een Operationeel overleg plaats tussen het Data Team en het ontwikkelteam. Hierin worden prioriteiten besproken en updates over en weer gedeeld.
 - Opdrachtgever vervult in de overleggen de rol van voorzitter;
 - Opdrachtnemer neemt deel aan de overleggen met medewerkers van gelijkwaardig niveau als de deelnemers vanuit Opdrachtgever;
 - Opdrachtnemer maakt bij de start van de overeenkomst aan Opdrachtgever kenbaar wie deze deelnemers zijn (in de vorm van profielen);
 - Aan het operationeel overleg dienen operationeel betrokken personen deel te nemen, op Tactisch en Strategisch niveau dienen de deelnemers voldoende mandaat te hebben om als gelijkwaardige gesprekspartner te fungeren op dat betreffende niveau.
 - Opdrachtnemer garandeert stabiliteit en continuïteit in de inzet van de deelnemers, wat betekent dat het verloop in de betreffende deelnemers tot een minimum wordt beperkt. Indien er toch sprake is van wisselingen in inzet van de betreffende deelnemers dient deze geleidelijk en naadloos plaats te vinden. Opdrachtnemer dient hiervoor te zorgen.
- Eis 3: Opdrachtgever schaaft de frequenties en de functionarissen op of af als een ontstane situatie daarom vraagt.
- Eis 4: De Opdrachtnemer is verantwoordelijk voor de schriftelijke vastlegging van de overleggen met de afgesproken acties en verbeterpunten en het (digitaal) aanleveren hiervan uiterlijk één week na het overleg. De vorm is in overleg met opdrachtgever.
- Eis 5: De escalatiestructuur op hoofdlijnen houdt het volgende in en geldt voor zowel de externe als de interne communicatiestructuur:
- Operationeel escaleert allereerst bij tactisch niveau en indien nodig (bijvoorbeeld bij afwezigheid) op strategisch niveau;
 - Tactisch niveau escaleert altijd op strategisch niveau en rapporteert over:

- Door opdrachtnemer te nemen en/of genomen maatregelen bij afwijking o.b.v. de gemaakte afspraken en SLA;
- Actueel overzicht van de gemaakte afspraken, besluiten en actiepunten.

Zowel de opdrachtgever als opdrachtnemer dienen zich aan deze structuur te houden.

Eis 6: Indien er een geschil is ontstaan dat voortvloeit uit of verband houdt met de Overeenkomst, zal in eerste instantie de escalatiestructuur worden aangehouden om het geschil op te lossen. Indien het niet lukt om het geschil op te lossen en het geschil op strategisch niveau is geëscaleerd wordt er binnen één (1) week bijeengekomen tijdens een vergadering.

2.2 Vergoeding (commerciële eisen)

- Eis 7: De overeengekomen prijzen en tarieven staan vast tot en met 31-12-2024. Bij de prijzen en tarieven worden geacht te zijn inbegrepen: salariskosten, reis- en verblijfskosten en alle met de uitvoering van de werkzaamheden samenhangende kosten, alsmede algemene kosten, winst en risico. De bedragen zijn exclusief omzetbelasting.
- Eis 8: Standaardbrieven van Opdrachtnemer waarin ‘algemene’ indexeringen worden aangekondigd aan alle klanten van Opdrachtnemer worden terzijde gelegd. Indien daarin staat opgenomen dat Opdrachtgever zich akkoord verklaart indien niet voor een bepaalde datum is gereageerd, zal dit geen gelding hebben op de overeenkomst die naar aanleiding van deze aanbesteding is gesloten.
- Eis 9: Opdrachtnemer zendt facturen voor de werkzaamheden onder vermelding van het in de nadere overeenkomst (opdrachtverlening) vermelde referentienummer aan factuurverwerking@vfpf.nl in pdf formaat.
- Eis 10: Betaling van een door Pf goedgekeurde factuur vindt uiterlijk 30 kalenderdagen na ontvangst van de factuur plaats.
- Eis 11: Het DWH-Saas-dienstenaanbod moet een transparant en flexibel pay-per-usemodel omvatten, waarbij kosten nauwkeurig worden toegerekend op basis van daadwerkelijk gebruik. Facturering moet plaatsvinden op basis van meetbare parameters, zoals dataopslag, dataverkeer en verwerkingstijd.
- Eis 12: De SLR over de gehele DWH-dienstverlening moeten gebruikers inzicht geven in hun verbruik en de bijbehorende kosten om zodanig kostenoverschrijdingen te voorkomen en de efficiëntie te verbeteren. De vorm en distributie van de SLR-rapportages is de verantwoordelijkheid van de

Opdrachtnemer. Opdrachtgever is bereid om eventueel aan te sluiten bij de standaard rapportagemethodiek van de Opdrachtnemer, mits deze voldoet aan de gestelde eisen zoals beschreven in het programma van eisen en de SLA. Deze rapportages hoeven niet geautomatiseerd beschikbaar gesteld te worden. De rapportages dienen met meerdere medewerkers bij de Opdrachtgever gedeeld te worden.

- Eis 13: De Opdrachtnemer moet heldere prijsstructuren en prijsgaranties bieden gedurende een afgesproken periode, om onverwachte kostenstijgingen te voorkomen. Ook moeten er flexibele mogelijkheden zijn om de dienst uit te breiden of te verkleinen, zodat de kosten kunnen worden aangepast aan de veranderende behoeften van onze organisatie. Opdrachtgever streeft naar een zo goed mogelijke en voorspelbare OPEX van de DWH-dienstverlening.

2.3 Service Level Agreement

VfPf en de Opdrachtnemer zullen gezamenlijk een Service Level Agreement (SLA) opstellen waarin specifieke details worden vastgelegd over:

- Beschrijving van de services;
- Beschikbaarheid en bereikbaarheid van het datawarehouse;
- SLR eisen met o.a. prestatie indicatoren en verbruikskosten;
- Minimale verwerkingscapaciteit;
- Gegevensintegriteit en beveiliging;
- Verantwoordingsrapportages i.h.k.v. privacy en security;
- Schaalbaarheid;
- Ondersteuning en onderhoud;
- Rapportage en monitoring;
- Datakwaliteit en data governance;
- Beëindiging en compensaties.

- Eis 14: Opdrachtnemer draagt aantoonbaar zorg voor steeds de nieuwste versies van de ISO-certificeringen (zie eis 16) te behalen.

- Eis 15: Opdrachtnemer is bereid en in staat om samen met VfPf de SLA op te stellen en zich te committeren aan de hierin gemaakte afspraken.

- Eis 16: Opdrachtnemer is voorafgaand aan de in beheername van de Datawarehousedienst verantwoordelijk voor het opstellen van de SLA en DAP (in samenspraak met VfPf). De SLA dient te voldoen aan de ISO27001-standaard en goedgekeurd te zijn door VfPf. Het betreft initieel de ISO27001-versie 2017 met dito beheersmaatregelen uit de ISO27002; Het DAP sluit aan op de (vastgestelde) VfPf-processen.
- Eis 17: Opdrachtnemer conformeert zich aan de overeen te komen SLA en DAP, rapporteert en stuurt op naleving van de overeengekomen afspraken en KPI's.
- Eis 18: Opdrachtgever ontvangt minimaal elke maand service-levelrapportages over de Datawarehousedienst in een duidelijke en herkenbare structuur conform de SLA. Hierin komt minimaal terug:
- De beschikbaarheid van het datawarehouse;
 - Het resourcegebruik en de financiële consequenties daarvan;
 - Statistieken over incidenten;
 - Statistieken over systeemprestaties;
 - Technische (security)-kwetsbaarheden en securitygebeurtenissen en/of -incidenten;
 - Minimaal (maar niet uitsluitend) de resultaten o.b.v. de gemaakte afspraken en SLA;
 - Door opdrachtnemer te nemen en/of genomen maatregelen bij afwijking o.b.v. de gemaakte afspraken en SLA;
 - Een actueel overzicht van de gemaakte afspraken, besluiten en actiepunten.

Opdrachtnemer dient per bovenstaande punten de betreffende parameters te rapporteren.

2.4 Beheer, ondersteuning en wijzigingen

2.4.1 Beschikbaarheid

Beschikbaarheid is de mate waarin 'het systeem' operationeel en toegankelijk is.

- Eis 19: De bijgewerkte inhoud van de datamarts (business store) moeten dagelijks minimaal tussen 7:30 uur en 23:00 uur beschikbaar zijn. Het verversen van de rapportages kan dagelijks om 7:30 uur starten. Dit staat los van eis 26 en 27 voor wat betreft incidentbeheer. VfPf heeft geen standbydiensten buiten kantooruren, met een uitzondering voor security-incidenten.
- Eis 20: De DWH-Saas dienst heeft een minimale beschikbaarheid van 99,5% op kantoordagen. De beschikbaarheid wordt per maand gemeten door het aantal minuten dat het systeem beschikbaar was te delen door de totale tijd in minuten, in het betreffende 'service window' (7.30-23.00 uur tijdens kantoordagen). Per maand dient de DWH Saas een beschikbaarheid te

hebben van 98%. Dit is zowel binnen als buiten het 'service window' en inclusief (gepland) onderhoud.

Eis 21: Alle meldingen (incidenten, wijzigingen, problemen etc.) worden vastgelegd in een servicemanagementsysteem van de Opdrachtnemer. Zowel Opdrachtnemer als Opdrachtgever kunnen hierin issues aanmaken en aanpassen. Opdrachtnemer en Opdrachtgever hebben toegang tot dit systeem waarbij de issues ook aangepast kunnen worden, zodat alle meldingen centraal geregistreerd zijn en hierover in de SLR gerapporteerd wordt.

De data in het datawarehouse wordt voor de meeste bronnen dagelijks (in de nacht) ververst. Nadat de datamarts zijn ververst, worden de rapportages geüpdatet. Dit laatste doet VfPf zelf en duurt ongeveer 30 minuten.

Eis 22: De data uit bronsystemen wordt in de nacht ververst. Het ETL-proces start op zijn vroegst om 23:00 uur en duurt maximaal tot 7:30 uur. Uitzonderingen zijn situaties waarin onverwachte wijzigingen in brongegevens leiden tot een verstoring van de gegevensverversing.

2.4.2 Herstelbaarheid

Herstelbaarheid is de mate waarin de software-oplossing(en), in geval van een onderbreking of bij een fout, de direct betrokken gegevens herstelt en het systeem in de gewenste staat terugbrengt.

Eis 23: De Opdrachtnemer garandeert een adequate back-upprocedure voor het datawarehouse conform de SLA.

Eis 24: Opdrachtnemer test periodiek de kwaliteit van de back-ups, de draaiboeken en procedures en de tijd die het kost om een back-up terug te zetten (recovery). Recoveryprocedures moeten minimaal op jaarbasis worden getest conform het 'Business Continuity Plan' (BCP) van Opdrachtgever. Opdrachtnemer rapporteert hierover naar Opdrachtgever en indien de testuitslag hiertoe aanleiding geeft, neemt Opdrachtnemer actie om de back-upprocedure te verbeteren.

De eisen m.b.t. de retentietijden zijn als volgt:

- Dagelijkse back-up dient 1 maand bewaard te worden in de GRS storage binnen de EU.
- Maandelijks back-up dient 12 maanden bewaard te worden op immutable GRS storage binnen de EU.
- Jaarlijkse back-up, dient 12 maanden bewaard te worden op immutable GRS storage binnen de EU.

Back-upprocedure dient onderdeel te zijn van het Business-continuityplan van de opdrachtgever.

Eis 25: Opdrachtnemer is in staat de prestaties van de Datawarehousedienst te rapporteren, analyseren en hier adequaat op te acteren, wanneer de situatie hierom vraagt. Ten aanzien van deze eis biedt de Opdrachtgever aan Opdrachtnemer de ruimte om een adequate invulling te geven aan het meten, analyseren en rapporteren over de prestaties van de DWH-dienstverlening. Een belangrijke prestatie-indicator is de doorlooptijd van het ETL-proces per bronsysteem en de verandering hiervan in de tijd.

2.4.3 Incidentbeheer

Incidentbeheer is de wijze waarop gehandeld wordt en omvat het tijdsbestek waarin het incident verholpen wordt. VfPf gebruikt drie prioriteiten bij incidenten met een daarbij behorende maximale oplostijd.

Eis 26: Meldingen komen binnen bij VfPf of haar werkplekleverancier en worden vervolgens uitgezet bij de relevante betrokken partijen. Deze afspraken worden in een Operationeel Leveranciers Overeenkomst (OLA) vastgelegd.

Eis 27: De Opdrachtnemer draagt zorg voor een incidentbeheerproces met minimaal de onderstaande prioriteiten en oplostijden.

Prioriteit	Impact	Responstijd	Oplostijd
1 Zeer hoog	Een incidentmelding met de prioriteit 'Zeer Hoog' geldt als er kritieke beperkingen in de dagelijkse bedrijfsprocessen zijn opgetreden. Definitie kritieke beperkingen: dit zijn de tijdkritische taken die ten gevolge van een volledige 'system shutdown' of een verstoring van een hoofdproces in de afgenomen producten niet uitgevoerd kunnen worden. De melding 'Prioriteit Zeer Hoog' vraagt onmiddellijke verwerking daar de verstoring kan resulteren in serieuze schade.	<1 uur	<1 werkdag
2 Hoog	Een incidentmelding met de prioriteit 'Hoog' geldt als er ernstige beperkingen in het dagelijkse bedrijfsproces zijn opgetreden. Benodigde taken kunnen niet worden uitgevoerd. Dit is veroorzaakt door een verstoring of uitvallen van een (deel)proces in de productieomgeving. De melding vraagt snelle verwerking daar het tot een serieuze onderbreking in het volledige productiesysteem kan leiden.	<4 uur	<2 werkdagen
3 Normaal	Een incidentmelding met de prioriteit 'Normaal' geldt als er beperkingen in het dagelijkse bedrijfsproces zijn	<2 werkdagen	<5 werkdagen

	opgetreden die als niet-kritisch worden beoordeeld. Dit is veroorzaakt door een verstoring of uitvallen van een deelproces in de productieomgeving.		
4 Laag	Een incidentmelding met de prioriteit 'Laag' geldt als er geen of nauwelijks beperkingen in het dagelijkse bedrijfsproces zijn opgetreden. Dit is veroorzaakt door een verstoring of uitvallen van een niet-dagelijks of weinig gebruikt deelproces in de productieomgeving. Klanten of medewerkers worden niet geraakt.	<5 werkdagen	O.b.v. 'Best effort'

2.4.4 Wijzigingsbeheer en Onderhoud

Opdrachtnemer richt een effectief en efficiënt proces in voor wijzigingsbeheer en onderhoud t.b.v. mogelijke maatwerkoplossingen die moeten worden gebouwd voor VfPf. Opdrachtnemer houdt zich daarbij minimaal aan de volgende eisen:

- Eis 28: Voor wijzigingen en nieuwe initiatieven committeert de Opdrachtnemer zich aan het VfPf-wijzigingsproces en de overeengekomen SLA en DAP.
- Eis 29: De Opdrachtnemer test elke release adequaat voordat deze beschikbaar wordt gesteld voor het uitvoeren van een GebruikersAcceptatieTest (GAT) door de Opdrachtgever. De Opdrachtnemer stelt, wanneer gevraagd, de testresultaten beschikbaar aan Opdrachtgever.
- Eis 30: Nieuwe releases worden geautomatiseerd middels scripts/installatiepakketten naar de test-, acceptatie- en productieomgeving uitgerold, zodat geen handmatige acties uitgevoerd dienen te worden om een nieuwe release naar een omgeving te brengen. De verwachting is dat Opdrachtnemer verantwoordelijk is voor het voortbrengingsproces (promotie in de OTAP-straat) / beheren van een CI/CD-straat.
- Eis 31: De Opdrachtnemer zorgt ervoor dat een rollback uitgevoerd kan worden als een nieuwe release niet zoals bedoeld functioneert en teruggedraaid moet worden.
- Eis 32: De Opdrachtnemer zorgt dat noodzakelijk technisch onderhoud en wijzigingen tijdig zijn afgestemd met de Opdrachtgever voorafgaand aan de uitvoering, conform vast te stellen SLA.

2.5 Prestaties en schaalbaarheid

Het VfPf Data Team bestaat momenteel uit acht personen welke allen tegelijkertijd query's kunnen afvuren op het datawarehouse. Over het algemeen zullen dit niet al te complexe query's zijn op de business store. De periodieke rapportages hebben een grotere databehoeft. Query's vanuit deze rapportages zijn groter en complexer.

Eis 33: De SaaS-oplossing heeft bij normaal gebruik bij het opvragen van gegevens en het genereren van rapporten responstijden van maximaal 3 seconden voor query's op één tabel en maximaal 30 seconden bij 10 tabellen. Onder 'normaal gebruik' vallen niet de periodes waarin de data in het datawarehouse wordt ververs (de 'nachtrun') en ook niet de periodes waarin de PowerBI-rapportages worden ververs. Het gaat hier om niet meer dan drie gebruikers die buiten voornoemde periodes tegelijkertijd query's afvuren op de database. De responstijden hebben betrekking op de ad-hoc-query's (door het Data Team) en tussentijdse verversingen van enkelvoudige rapportages tijdens kantooruren. Met responstijd wordt de uitvoering van de query bedoeld.

Eis 34: Het datawarehouse moet in staat zijn om meerdere query's tegelijkertijd te ondersteunen zonder merkbare vertragingen. Ga uit van 3 tot 8 query's tegelijkertijd.

VfPf voert geen constante performancedruk uit op de Datawarehousedienst. Dit betekent dat er in de cyclus van rapportageresultaten piekmomenten in belasting zullen zijn op de Datawarehousedienst.

De gegevensvolumes zullen verder toenemen, enerzijds omdat we historie opbouwen en anderzijds omdat het aantal gegevensbronnen zal toenemen. Tegelijkertijd zal het bevragen van de gegevens, zowel via standaardrapportages als via ad-hoc-query's toenemen. Indien bij groeiende volumes de prestaties afnemen, dan moet het systeem (de combinatie van applicatie en infrastructuur) kunnen worden opgeschaald.

Eis 35: Bij toenemende gegevensvolumes en gebruiksbelasting blijft het systeem schaalbaar en voldoen aan Eis 33 en 34.

Het Data Team van VfPf moet de mogelijkheid hebben om data via SQL-query's op te vragen (zie ook 3.2). In uitzonderlijke gevallen kan een verkeerd geschreven query zorgen voor een onnodige belasting van de database met de daarbij behorende kosten voor gebruik.

Opgevraagde query's via PowerBI zijn inzichtelijk door leverancier via beheer van Snowflake. De leverancier heeft in eerste instantie geen invloed op deze query's, maar via overleg en het bouwen van fail-safes kunnen query's met zeldzaam hoog verbruik worden gesignaleerd.

Eis 36: De Datawarehousedienst stopt automatisch query's die meer verwerkingscapaciteit vragen dan gebruikelijk. De criteria hiervoor worden in samenspraak tussen VfPf en de Opdrachtnemer bepaald.

VfPf wil graag inzicht in het gebruik van resources, zodat we kunnen bepalen waar eventuele kostenbesparingen nodig en mogelijk zijn.

Eis 37: VfPf ontvangt minimaal eenmaal per maand overzichten waarin duidelijk het resourcegebruik (bijvoorbeeld met credits voor gebruik en opslag) per gebruiker, per periode, per omgeving

wordt weergegeven. VfPf wil een zo goed mogelijke en voorspelbare OPEX. Hoe de Opdrachtnemer het verbruik inzichtelijk wil maken is leveranciersoplossing afhankelijk.

2.6 Informatiebeveiliging en privacy

- Eis 38: Opdrachtnemer beschikt over het ISO27001-certificaat en de door Opdrachtnemer te leveren dienst(verlening) en/of product (zoals omschreven in deze opdracht) is in scope van dit certificaat. Daarnaast heeft Opdrachtnemer passende beheersmaatregelen getroffen afkomstig uit de ISO27002, hetgeen blijkt uit de aan het certificaat gekoppelde verklaring van toepasselijkheid. Indien Opdrachtnemer gebruik maakt van een Cloud Service Provider (CSP), dan zijn tevens de ISO27017 en ISO27018 van toepassing op Opdrachtnemer dan wel de CSP die door Opdrachtnemer wordt ingeschakeld voor het uitvoeren van de dienstverlening.
- Eis 39: Opdrachtnemer draagt er aantoonbaar zorg voor steeds de nieuwste versies van de ISO-certificeringen te behalen. Opdrachtgever controleert dat Opdrachtnemer altijd over een geldig certificaat (incl. accurate Verklaring van Toepasselijkheid) beschikt. Het is daarmee aan Opdrachtnemer te zorgen dat certificaten niet verlopen en tijdig vervangen of herijkt worden.
- Eis 40: Bij de beveiliging van informatie wordt uitgegaan van de BIV-classificatie, te weten: beschikbaarheid (continuïteit), integriteit en vertrouwelijkheid. Opdrachtgever is verantwoordelijk voor deze BIV-classificatie. De Opdrachtnemer garandeert de beschikbaarheid, integriteit en vertrouwelijkheid van de systemen, diensten en de brongegevens die door Opdrachtgever wordt gehanteerd: Beschikbaarheid = Hoog, Integriteit = Hoog en Vertrouwelijkheid = Hoog. Opdrachtnemer voldoet aan de hieruit voortvloeiende in de Maatregelenmatrix van VfPf opgenomen eisen en maatregelen.
- Eis 41: Het datawarehouse is geen losstaande verwerking van persoonsgegevens, maar een ondersteunende applicatie voor de reguliere bedrijfsprocessen van VfPf, door te faciliteren in rapportages waarbij door middel van data-analyse specifieke (ethische) vraagstukken binnen het primair onderwijs in kengetallen tot uitdrukking worden gebracht. DWH sluit dan ook aan bij de verwerking van die bedrijfsprocessen en de daarbinnen welbepaalde, uitdrukkelijk omschreven en gerechtvaardigde doelen. Het uitgangspunt is dat DWH in beginsel geen persoonsgegevens verwerkt, echter om rapportages te genereren worden (persoons)gegevens uit andere bronnen opgehaald. In de huidige situatie kan niet worden voorkomen dat zich in de systemen van DWH persoonsgegevens bevinden; voor het maken van de rapportages die geen persoonsgegevens bevatten worden wel veelal toch ook door DWH persoonsgegevens verwerkt. Daar waar de DWH in aanraking komt met persoonsgegevens is Opdrachtgever verwerkingsverantwoordelijke en Opdrachtnemer verwerker in de zin van de AVG. Dit betekent dat Opdrachtnemer een verwerkersovereenkomst sluit, uitgaande van de versie van Opdrachtgever, zie bijlage 8.

- Eis 42: Opdrachtnemer is als verwerker verantwoordelijk en aansprakelijk voor alle subverwerkers die namens Verwerker ingezet worden om deze Opdracht te vervullen; aantoonbaar worden minimaal dezelfde eisen door Verwerker aan subverwerkers gesteld.
- Eis 43: Opdrachtnemer voldoet aan de voor het VfPf geldende wet- en regelgeving omtrent privacy en treft daartoe passende technische en organisatorische maatregelen, waarmee in ieder geval voldaan wordt aan de gangbare vormen van data-encryptie en beveiligingsprotocollen. De dienst is aantoonbaar ingericht dat zij Opdrachtgever ondersteunt en in staat stelt om te voldoen aan haar informatiebeveiligingsbeleid en alle relevante wet- en regelgeving. Partijen hebben de inspanningsverplichting elkaar hierover continu op de hoogte te houden.
- Eis 44: Opdrachtgever herhaalt – zie onder meer eis 41 – het uitgangspunt dat in beginsel in de ontwikkel-, test-, acceptatie- en productieomgeving geen persoonsgegevens worden gebruikt; het gebruik van persoonsgegevens wordt dus vermeden. In het geval dat de noodzaak bestaat toch persoonsgegevens te gebruiken in de productieomgeving - bijvoorbeeld omdat in de huidige situatie nog sprake is van verwerking van persoonsgegevens - worden alle persoonsgegevens geanonimiseerd, zodanig dat niet tot de persoon kan worden terug herleid. Anonimiseren betekent voor Opdrachtgever dat in alle gevallen gegevens nooit tot individuen herleid mogen/kunnen worden. Opdrachtgever definieert Persoonsgegevens conform de definitie in het eerste lid van artikel 4 van de AVG: 'alle informatie over een geïdentificeerde of identificeerbare natuurlijke persoon ("de betrokkene"); als identificeerbaar wordt beschouwd een natuurlijke persoon die direct of indirect kan worden geïdentificeerd, met name aan de hand van een identificator zoals een naam, een identificatienummer, locatiegegevens, een online identificator of van een of meer elementen die kenmerkend zijn voor de fysieke, fysiologische, genetische, psychische, economische, culturele of sociale identiteit van die natuurlijke persoon'.
- Eis 45: Opdrachtnemer hanteert – net als Opdrachtgever - het uitgangspunt dat er geen persoonsgegevens, noch bijzondere persoonsgegevens, in het DWH worden opgeslagen. DWH is geen eigen bronbestand/opslag; DWH haalt uit andere bronnen uitsluitend die gegevens op die noodzakelijk zijn voor het genereren van rapportages, analyses en voorspelmodellen. Indien DWH nog historische data bevat met persoonsgegevens, dan draagt Opdrachtnemer er zorg voor dat deze persoonsgegevens zo snel mogelijk worden verwijderd met dien verstande dat de oplossing de primaire processen niet mag verstoren. Het opslaan van historische data waarin geen persoonsgegevens zitten (geanonimiseerd dan wel onherkenbaar gemaakt tot de persoon) is wel toegestaan en Leverancier draagt zorg voor de veilige en integere opslag hiervan.
- Eis 46: Opdrachtnemer is zich uitdrukkelijk bewust van het volgende. Voor DWH wordt door Opdrachtgever per rapportage/voorspelmodel de bewaartermijn bepaald en geïmplementeerd, ook als er nieuwe rapportages/voorspelmodellen opgesteld worden. DWH bewaart uitsluitend die persoonsgegevens die voor een rapportage/voorspelmodel noodzakelijk zijn. DWH bewaart

deze gegevens niet langer dan nodig is voor het gebruik van die rapportage/dat voorspelmodel. Deze overbodige gegevens worden zo snel mogelijk verwijderd. Opdrachtnemer is verantwoordelijk dat de inrichting van bewaartermijnen - inclusief handhaving hiervan - geschiedt volgens de door Opdrachtgever gestelde bewaartermijnen.

- Eis 47: Opdrachtnemer draagt er zorg voor dat DWH controlemechanismes bevat om de juistheid van alle (persoons)gegevens die ten behoeve van de rapportage worden gegenereerd voortdurend te controleren voordat deze ingelezen worden in DWH.
- Eis 48: Opdrachtnemer hanteert de volgende principes. 'Privacy by design' en 'default' staan centraal bij de dienstverlening, en Opdrachtnemer hanteert deze waar mogelijk bij de verwerking en bescherming van persoonsgegevens encryptie c.q. versleuteling en pseudonimisering. Voor encryptie geldt als standaard de 'Handleiding Encryptiebeleid (PKI) voor privacy by design' versie 2.0 van maart 2019 van de Informatiebeveiligingsdienst. Voor privacy by design geldt als standaard de meest recente versie van de Handleiding Privacy Design van het CIP (Centrum Informatiebeveiliging en Privacybescherming). Opdrachtnemer voldoet - aantoonbaar aan Opdrachtgever – aan de vereisten die in beide documenten gesteld worden, richt deze eisen en uitgangspunten vanaf start van de Opdracht in en stemt op dit punt nader af met VfPf.
- Eis 49: Verwerker verwerkt alleen persoonsgegevens binnen de Europese Economische Ruimte (EER). Doorgifte aan landen hierbuiten is verboden. Hetzelfde geldt voor partijen die door Verwerker ten behoeve van vervulling van de opdracht worden ingeschakeld (subverwerkers).
- Eis 50: De afspraken rondom informatiebeveiliging en privacy worden vastgelegd in de SLA, inclusief prestatie-indicatoren en de bijbehorende verantwoordingsrapportages. Hierbij geldt dat de door het VfPf opgestelde DPIA (Data Protection Impact Assessment) integraal onderdeel is van deze dienst en Opdrachtnemer zich houdt aan alle daaruit voortvloeiende maatregelen dan wel nadere maatregelen die uit aankomende risicoanalyses naar voren komen. Onderkende risico's worden verwerkt in de vorm van aanvullende beveiligingsmaatregelen die door de Opdrachtnemer worden voorzien van de benodigde (borgings-)maatregelen ter mitigatie van de risico's.
- Eis 51: Strategisch Leveranciersmanagement Microsoft Rijk (SLM) heeft verschillende DPIA's uitgevoerd op de producten van Microsoft. Opdrachtnemer zal alle aanbevelingen, nu en in de toekomst van SLM doorvoeren. Dit betekent dat Opdrachtnemer aantoonbaar voldoet aan deze aanbevelingen en de bijbehorende acties doorvoert. Opdrachtnemer neemt het initiatief Opdrachtgever mee te nemen in ontwikkelingen en de voortgang van de voornoemde acties op dit punt.
- Eis 52: De Opdrachtnemer draagt zorg voor een back-upprocedure voor productiesystemen zoals door Opdrachtgever is beschreven in het Technisch securitybeleid en de bijbehorende

Maatregelenmatrix. Een consistente back-up moet kunnen worden teruggezet. Deze back-up moet periodiek getest worden. Met productiesystemen wordt binnen deze context bedoeld, alle systemen die nodig zijn waar Opdrachtnemer voor verantwoordelijk is om (eind)producten, zoals rapportages, te produceren.

- Eis 53: Binnen de scope van de dienstverlening vindt logging plaats, met als verplichte standaard de set met beheersmaatregelen uit hoofdstuk 12.4 van de ISO27002-norm; voor de verdere invulling en uitwerking van logging is de 'Handreiking Aanwijzing Logging versie 2.0' (maart 2019) van de Informatiebeveiligingsdienst verplicht. Dit omvat logging van bestanden die een audit-trail vastleggen in het kader van de hoofdstukken 16 en 17 van de ISO27002.
- Eis 54: Binnen de scope van de dienstverlening wordt toegang tot persoonsgegevens en overige gevoelige informatie afgeschermd, zodat alleen personen toegang hebben tot informatie die noodzakelijk is voor hun eigen werkzaamheden. De gebruikersaccounts zijn persoonsgebonden. Daarnaast wordt binnen het technisch beheer het volgende afgedwongen:
- Functiescheiding: fraudegevoelige processen liggen niet in de hand van één medewerker;
 - Verificatie: belangrijke administratieve handelingen moeten door een andere medewerker worden beoordeeld en geëffectueerd via het zogenaamde '4-ogen principe';
 - Toegang op basis van need-to-know;
 - De werkplek-tenant wordt gebruikt om de werkplekken van medewerkers van Opdrachtgever te beheren, de toegang van medewerkers te reguleren en diverse applicaties te beheren. Om aan de vereisten te voldoen, draagt Opdrachtnemer zorg voor een B2B-verbinding tussen de werkplek-tenant en de 'DWH'-tenant. In beide tenants is de andere tenant geregistreerd als nieuwe organisatie, zodat het uitwisselen van gegevens tussen beide tenants mogelijk is.
 - Wanneer gebruikers inloggen op de werkplek-tenant, worden ze geauthentiseerd middels multi-factor authenticatie. Via een conditional access policy wordt gecontroleerd of de gebruiker probeert in te loggen via een geregistreerd apparaat. Gebruikers die toegang willen krijgen tot de data-toepassingen krijgen een gast-account op de 'Azure AD Tenant' (AAD) van de DWH-tenant. Dit gast-account verwijst naar het AAD-account van de gebruiker op de werkplek-tenant. Toegang is gebaseerd op persoonsgebonden accounts en vereist multi-factor authenticatie. Deze of een vergelijkbare en gelijkwaardige oplossing moet van toepassing zijn.
- Eis 55: Opdrachtnemer moet een privacy- en een securityaanspreekpunt en -verantwoordelijke hebben die in overleg met Opdrachtgever testen, beoordelingen en evaluaties van de doeltreffendheid van de technische en organisatorische maatregelen ter beveiliging reviewt.
- Eis 56: De aanbieder van de DWH SaaS dient de gebruikte applicaties en tooling (applicatiestack) up-to-date te houden door tijdige patches en updates toe te passen. Daarnaast moet de aanbieder

volledige verantwoordelijkheid accepteren voor het identificeren en wegnemen van alle securityrisico's binnen een redelijke termijn na ontdekking, om de integriteit en beveiliging van het systeem te waarborgen. Deze termijn moet in overeenstemming zijn met industrienormen en best practices.

- Eis 57: Opdrachtnemer beschikt over een adequaat proces voor het vernietigen van (persoons)gegevens, in de breedste zin van het woord. Opdrachtgever kan aanvullende eisen stellen met betrekking tot het veilig verwijderen, transport en het beheer van data/informatie (inclusief verwijderbare media, indien van toepassing). De systemen zijn zo ingericht dat gegevens op basis van wet- en regelgeving op gezette momenten automatisch verwijderd worden. In geval dat bepaalde datadragers overbodig worden, dienen deze vernietigd te worden door een hiervoor gecertificeerde partij conform NIST SP 800-88 Rev.1.
- Eis 58: Opdrachtgever heeft het recht audits/assessments (waaronder in ieder geval jaarlijks een pentest) te laten uitvoeren op het gebied van informatiebeveiliging en privacy. Hieronder valt tevens minimaal eenmaal per jaar de eigen organisatie auditen op de volgende onderwerpen:
- Beveiliging
 - Kwaliteitssysteem
- De schriftelijke resultaten van de jaarlijkse audits/assessments worden overlegd en besproken met Opdrachtgever. Opdrachtnemer werkt volledig mee aan privacy en securityaudits/-assessments, in welke vorm en op welk moment dan ook. De kosten die hiermee gemoeid zijn komen voor rekening van Opdrachtgever.
- Eis 59: Opdrachtnemer meldt securitygebeurtenissen én security-incidenten conform het Incident response plan van VfPf onverwijld maar uiterlijk binnen 24 uur na constatering van een gebeurtenis/incident aan Opdrachtnemer, en verleent volledige medewerking en deelt alle beschikbare informatie over de gebeurtenis. In het geval van een security-incident moet de Opdrachtnemer redelijkerwijs maatregelen treffen om de mogelijk nadelige gevolgen te beperken voortkomend uit het incident.
- Eis 60: Opdrachtnemer zorgt voor een adequate beveiliging op alle niveaus (fysiek, netwerk, applicaties, firewall, IDS/IPS, ACL, IP-accesslist, etc.) en monitort dagelijkse kwetsbaarheden om zero-day-exploits te voorkomen. Dit betreft in ieder geval: 3
- Opdrachtgever dient alle niveaus van NCSC-beveiligingsadviezen (openbaar) op te volgen, voor zover deze betrekking hebben op systemen die relevant zijn voor het leveren van de dienstverlening. Hierbij dient een H /H melding binnen 1 week opgelost te worden. Ook dient over deze relevante NCSC-adviezen gerapporteerd te worden, waarbij ook de opvolgingsacties naar aanleiding van deze adviezen worden vermeld;

- Het blokkeren van bepaalde e-mails, zoals spam, phishing, e-mailvirussen en malware. Eventuele onterecht tegengehouden e-mails moeten door de Service Desk van Opdrachtnemer weer teruggezet kunnen worden;
- Het beveiligen van de internetverbinding tegen alle mogelijke soorten aanvallen gebeurt via IDS, IPS en continue monitoring, bewaking en proactief en reactief beheer bij aanvallen;
- Het voorkomen, detecteren, blokkeren en verwijderen van virussen;
- Het offline bewaren van back-ups ouder dan een week (i.v.m. mogelijke ransomware);
- Handelingen van systeemadministrators en systeemoperators worden vastgelegd en deze zijn niet aanpasbaar;
- Alle koppelingen met DWH worden beveiligd tegen de laatste stand der techniek, met als uitgangspunt de meest recente versie van ICT-beveiligingsrichtlijnen voor Transport Layer Security (TLS);
- Alle hiervoor benodigde apparatuur, licenties en dergelijke zijn de verantwoordelijkheid van Opdrachtnemer.

Eis 61: Er is een up-to-date lijst beschikbaar van de personen die bevoegd zijn om toegang te krijgen tot de systemen die worden ingezet voor het leveren van de gevraagde dienst. Deze lijst dient door Opdrachtnemer up-to-date gehouden te worden.

3 Functionele vereisten

3.1 Gegevensintegratie

Een van de belangrijkste functionaliteiten van de Datawarehousedienst is het ophalen, transformeren en verwerken van gegevens uit meerdere bronsystemen. Deze bronsystemen hebben hun eigen interfaces om data te ontsluiten. Veelvoorkomende interfacetypen zijn: REST (Json), SOAP, ODS (SQL), CSV-bestanden en Excel-bestanden. Zie paragraaf 1.4.2 Staging Area voor de huidige datakoppelingen.

De rolverdeling bij het aansluiten van nieuwe bronnen is in hoofdpunten als volgt:

- VfPf besluit welke bronnen gewenst zijn om aan te sluiten en wat hierin de prioriteit is.
- Samen met de leverancier van de Datawarehousedienst wordt bepaald welke selectie van gegevens dit betreft en hoe de koppeling technisch tot stand moet komen.
- De leverancier verzorgt vervolgens de technische koppeling en heeft waar nodig direct contact met de eigenaren van het bronsysteem.
- Doormodellering naar de Datavault en Datamarts gebeurt door de leverancier op basis van wensen en eisen vanuit VfPf.
- De leverancier test het geheel.

- VfPf doet zelf de acceptatietests.
- De leverancier verzorgt de deployment en de opname in de dagelijkse dataload.

Eis 62: De Datawarehousedienst moet in staat zijn om gegevens uit verschillende bronnen te integreren, waaronder interne systemen, externe gegevensbronnen en bestanden.

Eis 63: Het datawarehouse moet in staat zijn om een breed scala aan gegevensformaten te verwerken, zoals REST API (Json), SOAP, ODS (SQL), cloud services, CSV-bestanden, TXT-bestanden en Excel-bestanden.

Eis 64: De Datawarehousedienst moet de mogelijkheid bieden om gegevens te transformeren en te verrijken.

Eis 65: De geboden Datawarehousedienst heeft afdoende interfacemogelijkheden om zowel met oude als nieuwe leveringsprotocollen om te gaan.

Bronsystemen hebben elk hun eigen levenscyclus. Functionaliteiten en datastructuren kunnen veranderen en bronsystemen kunnen worden vervangen. Dit heeft consequenties voor de gegevensintegratie met het datawarehouse. In de meeste gevallen zullen wijzigingen worden aangekondigd. Zeker als het gaat om systeem migraties wordt het datawarehouse meegenomen in het project. Ook kleinere functionele wijzigingen in de bronsystemen kunnen consequenties hebben voor de datastructuur in de interface naar het datawarehouse. In uitzonderlijke gevallen worden deze niet van tevoren aangekondigd.

Eis 66: Het laadproces wordt gemonitord op afwijkingen in datastructuren. VfPf wordt uiterlijk de volgende werkdag op de hoogte gesteld van afwijkingen in dataformaten en -structuren. Zodra er afwijkingen zijn geconstateerd wordt verwacht dat de Opdrachtnemer actie onderneemt om het laadproces aan te passen. Wanneer hiervoor informatie vanuit VfPf nodig is, wordt hiervoor contact gezocht.

Eis 67: Het ETL-proces kan gemakkelijk worden aangepast aan veranderingen in bronstructuren en doelsystemen zonder dat dit grote wijzigingen in de code vereist.

Eis 68: In gevallen waar de datakwaliteit en reproduceerbaarheid (criteria te definiëren in samenspraak met VfPf) van het datawarehouse in gevaar komt wordt het laadproces gestaakt en teruggedraaid naar de oorspronkelijke staat.

VfPf wordt steeds meer datagedreven. Intensiever gebruik van data leidt tot meer vragen over de exacte herkomst en transformatie van gegevens in het datawarehouse. Het Data Team moet in staat zijn om deze vragen correct te beantwoorden. Het is niet uitgesloten dat VfPf op termijn software gaat gebruiken die data-lineage in kaart brengt.

- Eis 69: De ETL/ELT-processen zijn zorgvuldig gedocumenteerd. Van elk attribuut is bekend hoe deze tot stand is gekomen in termen van bronsystemen, transformaties en business rules. Deze documentatie is toegankelijk voor VfPf. In onze ogen is deze in de huidige situatie volledig en afdoende gedocumenteerd.
- Eis 70: In de Datawarehousedienst is een 'sandbox-omgeving' beschikbaar, waar de Data Analisten/Scientists en Data Specialist zelf (tijdelijk) data kunnen uploaden en wegschrijven naar tabellen. Deze data moet gekoppeld kunnen worden met de overige gegevens uit het datawarehouse, zowel de datamarts als de eerdere lagen, en dus beschikbaar zijn voor rapportage en analyse. Aan deze eis voldoet de huidige oplossing niet.

3.2 Gegevensanalyse en rapportage

VfPf gebruikt voornamelijk PowerBI en in mindere mate Excel voor rapportages. PowerBI wordt door VfPf beheerd en is geen onderdeel van deze aanbesteding. Het is van belang dat de bestaande PowerBI-rapportages met minimale inspanningen operationeel blijven. Het is voor VfPf geen optie om een andere rapportagetool dan PowerBI in te zetten.

Voor analyse en voorspelmodellen worden op dit moment Databricks, Excel en PowerBI ingezet. Het beheer van Excel en PowerBI valt niet onder deze aanbesteding, maar het is mogelijk dat voor Databricks door Opdrachtnemer een andere tool voorgesteld wordt. Hieronder zullen we de term 'data-scienceapplicatie' gebruiken voor de applicatie waarmee we analyses en voorspelmodellen maken en die wel onder deze aanbesteding valt. Deze term verwijst dus naar Databricks of mogelijk andere tooling met vergelijkbare functionaliteit.

Voor de rapportage- en analysetools die VfPf inzet, is een directe en beveiligde toegang tot het datawarehouse van belang.

De Data Analisten/Scientists en zeker de Data Specialist binnen VfPf maken ook veelvuldig gebruik van SQL-query's direct op de verschillende stores (Staging, Datavault en Datamart).

- Eis 71: De Opdrachtnemer biedt gedetailleerde documentatie en ondersteuning om ervoor te zorgen dat het VfPf Data Team effectief met het datawarehouse kan werken, ongeacht waar het wordt ingezet.
- Eis 72: De Datawarehousedienst heeft de mogelijkheid tot directe en veilige datakoppeling met SQL-clients, PowerBI (voor rapportages) en de data-scienceapplicatie (voor analyses en voorspelmodellen). Om data in deze applicaties te laden, hoeven geen losse (CSV-)bestanden te worden geëxporteerd en geïmporteerd.

- Eis 73: Het technisch beheer en de onderhoudstaken van de data-scienceapplicatie moeten in handen zijn van de Opdrachtnemer.
- Eis 74: In de data-scienceapplicatie moet het mogelijk te zijn om te werken met notebooks die Python, R en SQL ondersteunen als programmeertaal. De gebruiker moet hierin zelf de programmeertaal kunnen kiezen en verschillende programmeertalen in één notebook kunnen gebruiken.
- Eis 75: De data-scienceapplicatie moet de mogelijkheid bieden om (geautomatiseerd) tabellen vanuit de applicatie naar een sandbox-omgeving in het datawarehouse weg te schrijven. Aan deze eis voldoet de huidige oplossing niet.
- Eis 76: De data-scienceapplicatie moet de mogelijkheid bieden om naast data uit het datawarehouse ook gegevens uit externe bronnen, zoals losse CSV-bestanden en data uit webpagina's in te lezen en te gebruiken.
- Eis 77: De data-scienceapplicatie moet efficiënt en schaalbaar zijn om grote datasets te kunnen verwerken en complexe analyses uit te voeren; de data-scienceapplicatie moet goede prestaties leveren bij het uitvoeren van rekenintensieve taken, zoals machinaal leren en data mining, ook bij in de tijd groeiende datasets.
- Eis 78: De data-scienceapplicatie moet functionaliteit bieden waarmee geautomatiseerd en zonder te hoeven programmeren voorspelmodellen ontwikkeld kunnen worden. Dit betekent dat de data-scienceapplicatie in ieder geval aan de volgende punten moet voldoen:
- De data-scienceapplicatie moet bij het ontwikkelen van modellen de mogelijkheid bieden voor automatische datavoorbewerking, zoals ontbrekende waarden imputeren, normalisatie van gegevens en codering van categorische variabelen.
 - De data-scienceapplicatie moet automatisch het meest geschikte machine learning model kunnen selecteren op basis van de gegeven dataset en het doel van de gebruiker. Echter, de gebruiker moet hiervan kunnen afwijken en een model kunnen selecteren dat niet volgens het geautomatiseerde proces als 'beste' is geselecteerd.
 - De data-scienceapplicatie moet automatisch hyperparameter tuning kunnen uitvoeren. De applicatie optimaliseert de hyperparameters van de geselecteerde modellen om de prestaties te maximaliseren.
 - De data-scienceapplicatie moet gebruik maken van de schaalbaarheid van het platform en parallelle verwerking om de efficiëntie van modelontwikkeling te verbeteren, zelfs bij grote datasets.
 - De data-scienceapplicatie moet na modevaluatie automatisch een notebook opleveren met de broncode voor het voorspelmodel die, indien gewenst, aangepast en gereproduceerd kan worden.

- De data-scienceapplicatie moet modellen automatisch kunnen implementeren en beheren. Dit betekent dat modellen automatisch uitgescoord kunnen worden op nieuwe (al dan niet real-time) datasets. Hiervoor moet de data-scienceapplicatie automatisch een notebook opleveren dat, indien gewenst, aangepast kan worden. Zie paragraaf 1.4.6 voor nadere toelichting m.b.t. de huidige oplossing.

Eis 79: De datascience-applicatie moet een workflowmanagementfunctie bevatten waarmee VfPf taken kan automatiseren. Deze workflows moet VfPf zelf kunnen maken. Zie paragraaf 1.4.6 voor nadere toelichting m.b.t. de huidige oplossing.

Momenteel hebben de Databricks-gebruikers admin-rechten en kunnen zij Python-libraries op een cluster (Virtual Machine) installeren. Daarnaast kunnen gebruikers in een Databricks Notebook een library tijdelijk installeren voor de duur van de sessie. Deze installatie vindt niet plaats op de lokale computer en is een standaard functionaliteit van Databricks. Databricks heeft een veiligheidsbeleid om de risico's van malware en andere beveiligingsproblemen te beperken, zie ook [deze link](#).

Eis 80: Gebruikers moeten in staat zijn om packages en bibliotheken in de data-scienceapplicatie te installeren om aanvullende functionaliteiten toe te voegen. Het is vereist dat de data-scienceapplicatie een eenvoudig en intuïtief mechanisme biedt voor het installeren en beheren van externe packages.

Eis 81: De data-scienceapplicatie moet uitgebreide documentatie en handleidingen bevatten die gebruikers ondersteunen bij het gebruik van de functionaliteiten.

Eis 82: De Opdrachtnemer moet zorgen voor adequate technische ondersteuning voor de data-scienceapplicatie.

3.3 Gegevensbeheer

Voor VfPf is het van belang dat resultaten reproduceerbaar zijn. Indien een rapport is opgesteld voor een bepaalde tijdsperiode, en de brondata en business ruling is daarna ongewijzigd, dan moet dit rapport te allen tijde hetzelfde resultaat geven. Zonder uitdrukkelijk verzoek van VfPf of vastgelegde retentieprocedures wordt geen brondata verwijderd.

Eis 83: Resultaten vanuit dataquery's zijn eenvoudig reproduceerbaar.

Datakwaliteitsproblemen worden zoveel mogelijk opgelost binnen de bronsystemen. Het datawarehouse heeft hierin voornamelijk wel een signaleringsfunctie. In uitzonderlijke gevallen worden gegevens binnen het datawarehouse zelf opgeschoond, bijvoorbeeld wanneer een gegevensbron is uitgefaseerd.

VfPf heeft regels waaraan bepaalde gegevens moeten voldoen. Denk hierbij aan het voldoen aan dataformaten en referentiële integriteit. Datakwaliteitsregels worden geregeld aangepast en uitgebreid. Gegevens die niet voldoen aan de datakwaliteitsregels zijn duidelijk herkenbaar en rapporteerbaar. Momenteel is dit naar de huidige eisen voldoende ingericht, maar hierbij dient opgemerkt te worden dat wij datakwaliteitsprocessen wensen te intensiveren. Momenteel maken we nog geen gebruik van een specifiek model om de datakwaliteit op orde te houden.

- Eis 84: De Datawarehousedienst biedt mogelijkheden tot gegevensvalidatie en gegevensreiniging. Datakwaliteitsproblemen zijn herkenbaar, traceerbaar en eenvoudig te delen met de key users en functioneel coördinatoren van bronsystemen bij Opdrachtnemer.
- Eis 85: Opdrachtnemer en Opdrachtgever zijn bereid een OLA (Operational Level Agreement) af te sluiten waarin afspraken staan met betrekking tot de inspanningsverplichting om de datakwaliteit continu op een gewenst peil te houden en daar in een gezamenlijke inspanning met procedures en afspraken aan bij te dragen. Een OLA dient om de samenwerkingsafspraken tussen verschillende VfPf-leveranciers vast te leggen. Er worden door de Opdrachtgever geen eisen gesteld aan het gebruik van een bepaalde OLA-template. Van de opdrachtnemer wordt verwacht na te gaan of datakwaliteitsproblemen zijn ontstaan door processen onder verantwoordelijkheid van de Opdrachtnemer, bijvoorbeeld door het niet volledig inladen van de gegevens uit een databron of door technische problemen in de ETL-tool. Het oplossen van dit soort problemen valt ook onder de inspanningsverplichting. Voor datakwaliteitsproblemen die hun oorzaak vinden in de bronsystemen zelf, wordt alleen een signalering geëist.
- Eis 86: Opdrachtnemer is in staat datakwaliteitsproblemen te vertalen naar consequenties voor de business. Opdrachtnemer kan bij specifieke datakwaliteitsproblemen aangeven hoe deze problemen zich vertalen naar incorrecte rapportages, analyses en selecties.

3.4 Ongestructureerde data

VfPf heeft de ambitie om te kunnen groeien in haar data-analysebehoefte. Het gebruik van ongestructureerde teksten als data is hierin essentieel. Het is noodzakelijk dat de Datawarehousedienst mee kan groeien in deze ambitie. Er zijn vooralsnog geen plannen om data in de vorm van video, foto's of geluid op te slaan.

- Eis 87: De Datawarehousedienst heeft de mogelijkheid om ongestructureerde tekstdata op te slaan en beschikbaar te maken voor analyse.
- Eis 88: De data-scienceapplicatie moet functionaliteiten bevatten om ongestructureerde tekstdata te analyseren (text mining).

4 Technische vereisten

Portabiliteit

- Eis 89: Het datawarehouse moet met beperkte aanpassingen platformonafhankelijk kunnen draaien op verschillende besturingssystemen en hardware-infrastructuur. Met platformonafhankelijk wordt bedoeld dat Opdrachtgever zo min mogelijk maatwerk wil en zo veel als mogelijk van de standaarden gebruik wil maken. Dit betekent o.a. dat het mogelijk moet zijn om het datawarehouse gemakkelijk te migreren van het ene platform naar het andere zonder aanzienlijke aanpassingen aan de code of infrastructuur.
- Eis 90: Het datawarehouse moet compatibel zijn met meerdere DBMS-en waardoor het gemakkelijk overgezet kan worden naar verschillende databaseplatforms.
- Eis 91: Het moet mogelijk zijn om het datawarehouse gemakkelijk te migreren zonder aanzienlijke aanpassingen aan de code of infrastructuur naar meer marktconforme platformen indien noodzakelijk.
- Eis 92: Het datawarehouse moet uitgebreide instrumentatiemogelijkheden hebben om de prestaties, beschikbaarheid en gezondheid van het systeem te monitoren en te beheren, ongeacht het platform waarop het wordt uitgevoerd.

Compatibiliteit

- Eis 93: Het datawarehouse moet haar functionaliteiten efficiënt kunnen uitvoeren in een omgeving waarin meerdere producten bestaan. Dit, zonder nadelige invloed op de andere producten.

Schaalbaarheid

- Eis 94: Het datawarehouse moet in staat zijn om te schalen door het toevoegen van meer hardwareressources, zoals CPU, geheugen en opslag, om de prestaties te verbeteren en aan de groeiende behoeften te voldoen.
- Eis 95: Het datawarehouse moet monitoringtools en -rapporten bieden waarmee u de schaalbaarheid kunt volgen, belasting kunt meten en proactieve beslissingen kunt nemen over het opschalen van resources. Het uitgangspunt is dat de Opdrachtnemer eigenaar is en blijft van de benodigde licenties.

Logging

- Eis 96: Activiteiten van gebruikers, uitzonderingen en informatiebeveiligingsgebeurtenissen behoren te worden vastgelegd in audit-logbestanden. Deze logbestanden behoren gedurende een periode van 12 maanden te worden bewaard, ten behoeve van toekomstig onderzoek en toegangscontrole. Bij een (vermoedelijk) informatiebeveiligingsincident is de bewaartermijn minimaal drie jaar.
- Eis 97: Logberichten mogen alleen inzichtelijk zijn voor specifieke geautoriseerde personen.
- Eis 98: Het wegschrijven van logberichten mag geen invloed hebben op de performance van de applicaties.
- Eis 99: Logbestanden en de instellingen van logmechanismen worden zodanig beschermd dat deze niet aangepast of gemanipuleerd kunnen worden. Indien de instellingen aangepast moeten worden zal daarbij altijd het '4-ogen principe' toegepast worden.

Technische ondersteuning en onderhoud

- Eis 100: De Opdrachtnemer moet technische ondersteuning bieden voor eventuele systeemp Problemen of vragen.
- Eis 101: Het systeem moet regelmatig worden onderhouden en bijgewerkt en altijd voldoen aan de nieuwste beveiligings- updates en patches.
- Eis 102: Er ligt een duidelijk plan voor software-updates en onderhoud van het datawarehouse waar rekening is gehouden met een minimale downtime tijdens de updates en patches zoveel mogelijk buiten kantooruren.

4.1 Hosting en infrastructuur

- Eis 103: De datacenterfacility-componenten (UPS, Cooling, energieleveringspaden) zijn alle redundant uitgevoerd (Tier 3 datacenter) en daardoor gelijktijdig onderhoudbaar.
- Eis 104: Het systeem moet worden gehost in datacenters die voldoen aan de hoogste normen voor beveiliging en beschikbaarheid.
- Eis 105: De Opdrachtnemer moet zorgen voor een betrouwbare hostinginfrastructuur met voldoende capaciteit en schaalbaarheid.
- Eis 106: De Opdrachtnemer is verantwoordelijk dat op de infrastructuur altijd support wordt gegeven door bij te zijn met de laatste software- en middleware-updates. Bij grote updates waarbij de

dienst langere tijd niet beschikbaar zal zijn, wordt in overleg met VfPf het juiste tijdstip bepaald om de update door te voeren en te testen.

Ondersteuning en onderhoud

- Eis 107: De Datawarehousedienst dient afzonderlijk van productie ook aangesloten en functioneel te zijn op de acceptatieomgeving van het systeem t.b.v. test, doorontwikkelingen en training. VfPf verwacht het voortbrengingsproces (release pipeline)/ CI/CD-straat van de Opdrachtnemer, mits van toepassing.
- Eis 108: De Opdrachtnemer zorgt ervoor dat de acceptatie- en productieomgeving van het datawarehouse functioneel identiek aan elkaar zijn.
- Eis 109: Alle releases (SaaS-oplossingen) volgen het VfPf-wijzigingenproces waarbij elke release formeel door de Opdrachtgever wordt vrijgegeven voordat deze naar de productieomgeving wordt doorgezet. Opdrachtgever wil in control zijn over de productieomgeving en zal aanpassingen daarop volgens het wijzigingenproces laten verlopen. Waar dit niet mogelijk is, wil Opdrachtgever hier tijdig vooraf over geïnformeerd worden.

5 Eisen ten aanzien van continuering

Voor de continuering van de Datawarehousedienst zijn meerdere scenario's mogelijk. Kennis en ervaring van de tools bij de leverancier, een korte overgangsperiode, langetermijn-beheersbaarheid en groeipotentie zijn voor VfPf belangrijk. VfPf heeft een voorkeur voor standaardsoftware boven maatwerk, maar de uiteindelijke keuze voor de tooling en technieken in het voorstel van de Opdrachtnemer is aan de Opdrachtnemer zelf.

Snowflake is ANSI SQL99-compliant. Het biedt een exportfunctie voor migratie van de database. De huidige Opdrachtnemer kan de nieuwe partij ondersteunen bij de migratie. De migratie van de ETL/ELT-processen is sterk afhankelijk van de functionaliteiten van Matillion en de nieuwe tool. VfPf heeft onvoldoende specifieke kennis om dit te beoordelen. Het is mogelijk om van modelleringmethoden als Datavault 2.0 af te wijken mocht de Opdrachtnemer van mening zijn dat dit beter bij de eisen en wensen van VfPf past.

We verwijzen naar het implementatieplan (wens 1) voor een duidelijke beschrijving van het transitiescenario.

In een scenario waarin de Microsoft-tenant met de huidige tooling (Snowflake, Matillion en Databricks) één-op-één wordt overgenomen, heeft de huidige Opdrachtnemer ten aanzien van de kennisoverdracht de onderstaande inspanning ingeschat. Mocht hier meer tijd voor nodig zijn dan kan dit besproken worden.

- Technische keuzes en ontwerpbeslissingen: 4 uur;
- Databronnen: 2 uur per bron;
- Data dictionary: 1 uur;

- Datamarts: 4 uur.

De datakoppelingen zijn in detail gedocumenteerd, van staging naar datavault en van datavault naar datamart. Bij overgang naar een nieuwe Opdrachtnemer draagt de huidige Opdrachtnemer de documentatie over. Wanneer van toepassing stelt de huidige Opdrachtnemer een export van de laatste versie van de Datawarehousedienst vanuit de GIT-repository ter beschikking.

Voor VfPf is het van belang dat de ontwikkelingen en data-ambities niet stil komen te liggen. We willen als VfPf zo weinig mogelijk groeibelemmeringen hebben tijdens de overgangperiode. We zijn ons ervan bewust dat het een complex systeem betreft waar zorgvuldigheid en betrouwbaarheid prevaleren boven snelheid. We verwachten desalniettemin van de nieuwe Opdrachtnemer een inspanning om de periode waarin VfPf geen vooruitgang op BI-vlak kan boeken, zo kort mogelijk te houden.

- Eis 110: Opdrachtnemer dient binnen vier weken na de definitieve gunning van de dienstverlening een implementatieplan op te stellen en deze ter beoordeling aan Opdrachtgever voor te leggen. Opdrachtgever dient akkoord te gaan met dit implementatieplan vóór start van de implementatie. Opdrachtnemer doet er alles aan om het implementatieplan (in de basis) op orde te hebben binnen de gestelde vier weken na gunning. Temeer de beoogde Opdrachtnemer in de beantwoording van deze aanbesteding reeds een conceptueel implementatieplan dient op te leveren als beoordelingscriterium en kennis heeft van de gestelde kaders.
- Eis 111: Om de overdracht in goede banen te leiden, stelt Opdrachtnemer iemand aan die verantwoordelijk is voor alle activiteiten die tijdens de implementatie- en transitieperiode plaatsvinden.
- Eis 112: Op 1 januari 2025 is de Datawarehousedienst operationeel. Alle databronnen zijn aangesloten en gegevens worden bijgewerkt volgens de laadfrequentie in paragraaf 1.4.1. De datamarts zijn ververst met de meest recente gegevens. Alle bestaande rapportages, analyses en voorspelmodellen kunnen worden gebruikt.
- Eis 113: Op 1 januari 2025 beheert de nieuwe Opdrachtnemer de Datawarehousedienst op basis van eigen kennis. Er is voor het juist functioneren en beheren van de dienst geen afhankelijkheidsrelatie met ex-opdrachtnemers. De nieuwe Opdrachtnemer is verantwoordelijk voor de planning en de juiste informatie-uitvraag, ofwel de juiste overdracht van deze kennis. De oude Opdrachtnemer is hierin volgend. Tot aan het formele 'handover'-moment (bepaald in het implementatieplan) is de huidige Opdrachtnemer verantwoordelijk voor de continuering van de dienstverlening.
- Eis 114: Voorafgaand aan de inproductiename organiseert de Opdrachtnemer een formeel overdrachtsmoment aan RUN, waarbij projectdocumentatie (incl. actie- en besluitenlijst van het project) beschikbaar wordt gesteld conform de in-beheername-procedure van VfPf.

6 Doorontwikkeling datawarehouse

Data neemt een steeds belangrijkere plaats in binnen VfPf en de onderwijssector. VfPf heeft de ambitie om de rol van kenniscentrum binnen de sector Primair Onderwijs op zich te nemen. De ontwikkelingen op het datawarehouse staan daarom nooit stil. We richten ons voornamelijk op:

- a. Toevoeging van meer interne en externe gegevens aan het datawarehouse;
- b. De gegevensbetrouwbaarheid te vergroten;
- c. De toepasbaarheid van de gegevens te vergroten;
- d. Koppeling met VfPf-portalen.

Ad a. Toevoeging van gegevens

Er is op dit moment geen concreet uitgeschreven roadmap. We denken aan de volgende databronnen:

- Externe bronnen die voorspelbare waarde voor het verzuim hebben. Te denken valt aan diverse data van het CBS en additionele data vanuit DUO;
- Master-klantgegevens uit CRM, ter vervanging van DUO Brin-bron;
- Uitbreiding CRM-gegevens ten behoeve van betere effectmeting van onze diensten;
- Verdere doormodellering van Arbomeestergegevens. We richten ons nu voornamelijk op het responsproces. We willen ook graag de antwoorden op inhoud analyseren.

De aan te sluiten externe bronnen en de volgordelijkheid hiervan zijn afhankelijk van onder andere de implementatie van VfPf als kenniscentrum, de rapportagewensen uit de regiotteams en de gewenste stuurinformatie voor het managementteam. Deze zijn op dit moment nog niet concreet genoeg om te vertalen naar een roadmap. In de loop van 2024 zullen we deze roadmap gezamenlijk (ook met de Opdrachtnemer van de Datawarehousedienst) opstellen.

Onze eigen inschatting is dat wij 70% van de gewenste interne en externe bronnen al hebben aangesloten.

Ad b. Gegevensbetrouwbaarheid vergroten

Buiten de wettelijke taken van VfPf die een nauwkeurige datakwaliteit vereisen, komen er vaker en meer ophelderingsverzoeken vanuit de eindgebruikers over bijvoorbeeld verschillen tussen twee bronsystemen met daarbij de vraag welke nu de waarheid is. Er is echter nog geen concrete roadmap voor datakwaliteit of een kwantificering van de gewenste verbeteringen. Deze staat wel op de agenda van 2024 en verder. Belangrijk om te melden is dat datakwaliteit in eerste instantie een eis voor de bronsystemen is. Datakwaliteitsissues moeten daar worden opgelost. Het datawarehouse heeft voornamelijk een signaleringsfunctie, die wij willen uitbreiden en verbeteren.

Ad c. Toepasbaarheid van gegevens vergroten

Onze groeiambities hebben ook betrekking op de uitbreiding van het aantal afnemers van informatieproducten. Denk aan toegang tot een deel van onze rapportages voor schoolbesturen, werknemers- en werkgeversorganisaties binnen het PO, de PO-raad, OCW en onze eigen bestuur/auditcommissie. Binnen de sector primair onderwijs wordt data steeds meer omarmd. Dit betekent dat partijen steeds meer data vergaren en beschikbaar stellen. VfPf heeft vanuit zijn wettelijke taak en dienstverlening unieke data die binnen de sector gedeeld kan worden.

Ad d. Koppeling met portalen

VfPf wil in de toekomst interfaces tussen enerzijds het datawarehouse en/of PowerBI en anderzijds de VfPf-portalen. Binnen bijvoorbeeld het project 'Klantbeeld' zal VfPf voor zijn medewerkers steeds meer klantgegevens vanuit meerdere interne en externe systemen gaan ontsluiten naar een webportaal. Het datawarehouse zal naar alle waarschijnlijkheid één van deze interne bronnen zijn. Het portaal zelf wordt door een externe leverancier van VfPf ontwikkeld.

Er zijn twee mogelijke interfaces:

- Een directe datakoppeling tussen een (front-end)portaal en Snowflake, waarbij samenwerking met de Opdrachtnemer vereist is.
- Het 'embedden' van een PowerBI-rapportage in het portaal, waarvoor geen directe samenwerking met Opdrachtnemer voor nodig is.

Naast bovengenoemde drie onderwerpen richten we ons ook op de volgende verbeteringen:

- De technologische ontwikkelingen op BI-gebied (nieuwe tools, nieuwe werkwijzen) die bijdragen aan een verdere professionalisering van het Data Team en die onze rol als kenniscentrum kunnen verstevigen;
- Datamarts moeten meer een weergave worden van de businessprocessen en minder een weergave van de keuzes die in bronsystemen zijn gemaakt. De datamartstructuur willen we aansluiten op het (nog te ontwikkelen) Enterprise Data Model;
- Rekenregels die nu in PowerBI (DAX) worden toegepast dienen te worden verplaatst naar de Datavault/Datamart;
- Beschikbaarheid van een 'sandbox-omgeving' en de mogelijkheid tot wegschrijven van data in deze omgeving door de data-scienceapplicatie;

Eis 115: De Opdrachtnemer van de Datawarehousedienst is in staat om de groei-ambities ten aanzien van business intelligence te volgen, waar nodig VfPf te informeren over nieuwe ontwikkelingen en VfPf hierin te begeleiden. Tevens kan de Opdrachtnemer hierbij VfPf adviseren in de wijze van vormgeven van de business intelligence.

Eis 116: De Datawarehousedienst dient in de toekomst een interface mogelijk te maken met portalen van VfPf, om specifieke data vanuit het DWH te kunnen tonen. Aan deze eis voldoet de huidige oplossing niet.

In de operationele fase zullen nieuwe databronnen worden toegevoegd en bestaande databronnen worden vervangen. In deze gevallen houdt VfPf gebruikelijk de lijnen kort. Waar mogelijk wordt direct geschakeld tussen – vooral externe – leveranciers van de databron en de Opdrachtnemer van de Datawarehousedienst, zeker als het gaat om de optimale technische koppeling van beide systemen.

Eis 117: De Opdrachtnemer van de Datawarehousedienst is in staat om in nauwe samenwerking met de leveranciers van bronsystemen van VfPf nieuwe databronnen toe te voegen en bestaande databronnen te vervangen.

Eis 118: De Opdrachtnemer van de Datawarehousedienst is in staat om op basis van de specificaties van de datakoppeling en de gebruikerswensen van het Data Team een gedegen planning op te stellen voor het koppelen van nieuwe databronnen of vervangen van bestaande databronnen.

Een groeiende datagedrevenheid van VfPf leidt tot verdiepingen en verbredingen van datagerelateerde vragen. Er kunnen situaties zijn waarin de gegevens – op het juiste niveau – beschikbaar zijn in de staging store of modelled store, maar nog niet in de datamarts. Dan is het zaak om de datamarts hierop aan te passen.

Eis 119: De Opdrachtnemer van de Datawarehousedienst is in staat de business-intelligencebehoefte van VfPf te vertalen naar nieuwe en waar nodig verbeterde datamodellering van de datamarts.

Elke wijziging in de datawarehouseomgeving moet door de Opdrachtgever worden getest en daarna worden geaccepteerd. Over het algemeen zijn de acceptatiecriteria van tevoren bekend gemaakt. De acceptatie gebeurt op de acceptatieomgeving. Pas na acceptatie door VfPf wordt de wijziging doorgevoerd op productie.

VfPf hanteert onderstaande 'definition of done':

- De ontwikkeling is opgeleverd in een voor VfPf toegankelijke acceptatieomgeving;
- De ontwikkeling is getest door het ontwikkelteam;
- VfPf heeft het testplan en de testresultaten ontvangen;
- De documentatie is bijgewerkt en aan VfPf ter beschikking gesteld;
- De ontwikkeling is 'non-breakable' (een rollback is mogelijk);
- Er is voldoende representatieve data in de acceptatieomgeving beschikbaar om acceptatie te testen.

Eis 120: Nieuwe ontwikkelingen voldoen volledig aan bovenstaande 'definition of done'.

Eis 121: Het ontwikkelteam stelt het testplan en de testresultaten beschikbaar aan VfPf ter beoordeling.

Een team kan van samenstelling wijzigen. Wanneer nieuwe personen op bestaande rollen komen, is het van belang dat de inwerktijd minimaal is en de kennisoverdracht – zowel van tools/applicaties als VfPf-specifieke kennis – volledig. Van de Opdrachtnemer wordt verwacht dat hij hiervoor zorgdraagt.

Eis 122: De Opdrachtnemer levert de resources voor een ontwikkelteam en zorgt voor de continuering van de werkzaamheden, ook wanneer de samenstelling van het ontwikkelteam wijzigt. Opdrachtgever ziet de kennisborging als de verantwoordelijkheid van de Opdrachtnemer.

7 Exit criteria

Mocht er op termijn een einde komen aan het gebruik van de Datawarehousedienst door VfPf, dan is het goed om een beeld te krijgen wat dit betekent voor de overgangsperiode. Dit in het licht van kosten en snelheid, in combinatie met instandhouding van productie-servicelevels tijdens de overgang.

Eis 123: De Datawarehousedienst is eenvoudig overdraagbaar. Er is geen 'lock-in' bij de Opdrachtnemer voor wat betreft de infrastructuur of gebruikte software.

Eis 124: Om het einde van de opdracht (afloop contract of bij vroegtijdige beëindiging) in goede banen te leiden, wordt Opdrachtnemer geacht alle medewerking te verlenen die nodig is voor een geruisloze overgang. Op deze manier worden risico's (onder andere verlies van informatie en communicatie) vermeden en worden de activiteiten gerelateerd aan het overgangsproces effectief gemanaged.

Eis 125: Tijdens het overgangsproces is Opdrachtnemer verplicht om te blijven voldoen aan de gestelde eisen conform dit Programma van Eisen. De reeds geplande periodieke werkzaamheden dienen tot de einddatum uitgevoerd te worden. De reeds geplande opdrachten alsmede de resterende facturen worden afgehandeld. Hier dienen Opdrachtnemer en Opdrachtgever overeenstemming over te bereiken.

Eis 126: Opdrachtnemer is verantwoordelijk voor het opleveren van een exitplan dat ter toetsing wordt voorgelegd aan Opdrachtgever, 6 maanden na inproductie van het datawarehouse.

Dit exitplan dient tenminste de volgende onderwerpen te bevatten:

- eigenaarschap en overdracht van software, licenties, intellectuele eigendomsrechten, overeenkomsten met derden en medewerkers;
- de wijze van overdracht van kennis en relevante product- en projectdocumentatie die noodzakelijk is bij een re-transitie van de dienstverlening;
- projectafspraken over exit en re-transitie m.b.t. regie, planning, doorlooptijd en

concrete projectdeliverables;

- afspraken over de kosten en tarieven die doorbelast worden bij een exit en re-transitie;
- afspraken over een controle (audit) waarmee kan worden vastgesteld dat alle aan VfPf behorende eigendommen en data/informatie veilig is overgedragen en bij de Opdrachtnemer is vernietigd mede conform de eisen uit de afgesloten VfPf Verwerkersovereenkomst.

VfPf baseert haar uitgangspunten voor het exitplan onder andere op de 'Gedragsregels Re-transitie' van het Platform Sourcing Nederland.