



Belastingdienst

Marktconsultatie

Data Storage Platform (DSP)

Opdrachtgever: De Belastingdienst

Versie : 1.0

Datum : 17 mei 2023

INHOUDSOPGAVE

Hoofdstuk 1. Algemeen	3
1.1. De Belastingdienst	3
1.2. Doelstelling marktconsultatie	3
1.3. Randvoorwaarden marktconsultatie	3
Hoofdstuk 2. Marktconsultatie procedure	5
2.1. Algemeen	5
2.2. Contact	5
2.3. Marktconsultatiedocument	5
2.4. Planning marktconsultatie	5
2.5. Vormvereisten	5
2.6. Randvoorwaarden deelname aan deze marktconsultatie	6
Hoofdstuk 3. Context	7
3.1. Context	7
3.2. Beoogde Scope	7
3.3. Architectuur Principes	10
Bijlage 1: Begrippenlijst	12

Hoofdstuk 1. Algemeen

1.1. De Belastingdienst

De Belastingdienst maakt deel uit van het Ministerie van Financiën. Het ministerie van Financiën is verantwoordelijk voor het financieel-economisch beleid in Nederland, het beheer van de overheidsfinanciën en het beleid met betrekking tot de financiële markten. Het ministerie heeft een centrale rol bij het opstellen van de Rijksbegroting en de miljoenennota, en bewaakt de begrotingsuitgaven.

De bekendste taak van de Belastingdienst is het heffen en innen van belastingen en premies. Jaarlijks verwerkt de Belastingdienst de verschillende aangiften van ruim 6 miljoen particulieren en ongeveer 1,1 miljoen ondernemers. De Belastingdienst int niet alleen, maar keert ook uit.

Andere belangrijke taken zijn:

- Het opsporen van fiscale, economische en financiële fraude;
- Het houden van toezicht op de invoer, uitvoer en doorvoer van goederen;
- Toezicht houden op het naleven van de fiscale wetten en regels.

1.1.1. Inkoopuitvoeringscentrum (IUC) Belastingdienst

Het IUC Belastingdienst is onderdeel van Belastingdienst SSO facilitaire dienstverlening (B/SSOFD) en verricht generieke inkoopactiviteiten voor de bij haar ondergebrachte Rijkscategorieën. Daarnaast zorgt zij voor de specifieke inkoop van de Belastingdienst en Domeinen Roerende Zaken. Zie voor meer informatie over de organisatie van de Belastingdienst de internetsites www.belastingdienst.nl en www.rijksoverheid.nl/ministeries/fin. Het IUC Belastingdienst zal deze marktconsultatie begeleiden.

1.2. Doelstelling marktconsultatie

De Belastingdienst is voornemens een aanbesteding uit te voeren voor een Data Storage Platform (DSP). Met deze marktconsultatie willen wij verifiëren of de door ons opgestelde scope (zie hoofdstuk 4) aansluit bij wat de markt kan bieden. Hiertoe hebben wij het verzoek of u uw mening en visie op de door ons opgestelde scope en architectuurprincipes (hoofdstuk 3) kunt geven.

Wat vindt u van de voorgestelde scope en zou u hier een oplossing voor kunnen leveren?

Zou u de volgende onderwerpen mee kunnen nemen in uw beantwoording:

- Architectuur en topologie van de oplossing;
 - Welke onderdelen komen in aanmerking voor hybride of cloud deployment en waarom?
- Gebruik van open source en open standaarden;
- Cloudvisie (met name in de context van de vraag);
- Eventuele afhankelijkheden met technologie van andere leveranciers;
- Schaalbaarheid van de oplossing;
- Contractvormen, Licentiemodel & pricing;
- Integratie met koppervlakken (o.a. data catalog software);
- Het functioneel en technisch beheer van de oplossing;
- Vervanging hardware na afschrijving;
- Exit strategie en toekomstige transitie werkwijze;
- De wijze van aanbesteden (aanbestedingsprocedure(s)).

Mist u functionaliteiten die nu niet in de scope zijn opgenomen en kunt u aangeven waarom deze functionaliteiten in uw ogen relevant zijn.

Na ontvangst van de schriftelijke beantwoording willen wij u graag uitnodigen voor een dialoog over de gegeven antwoorden.

1.3. Randvoorwaarden marktconsultatie

Deze marktconsultatie dient uitdrukkelijk niet om een voorselectie te maken van gegadigden in het kader van een uit te voeren verwerving.

Marktpartijen komen door deel te nemen aan de marktverkenning niet in een bevoordeelde of benadeelde positie op het moment van een toekomstig verwervingstraject.

Deelnemers aan de marktconsultatie kunnen geen rechten ontlenen aan deelname aan de marktconsultatie of aan enige informatie die tijdens de marktconsultatie wordt verstrekt. Het honoreren van claims van marktpartijen over het gebruik van informatie of vertrouwelijkheid is niet mogelijk.

De Belastingdienst zal de informatie uit de marktconsultatie zo nodig gebruiken om de aanbesteding kwalitatief beter te maken. De Belastingdienst behoudt zich alle rechten voor bij de vaststelling van het uiteindelijke eventuele verwervingstraject en de daarbij horende aanbestedingsstukken.

Hoofdstuk 2. Marktconsultatie procedure

2.1. Algemeen



Deze marktconsultatie is aan meerdere partijen verstuurd. Wij ontvangen uw bijdrage aan de marktconsultatie graag zo spoedig mogelijk maar uiterlijk op de in paragraaf 2.4. vermelde datum.

2.2. Contact

De Belastingdienst hanteert voor deze formele marktconsultatie heldere regels, daarom dient u alleen met onderstaande contactpersonen over deze marktconsultatie te communiceren via het onderstaand e-mailadres.

Contactpersoon	
Naam:	R.R. (Richard) Kalksma
Functie:	Senior Inkoopadviseur IT
E-mail:	rr.kalksma@belastingdienst.nl

2.3. Marktconsultatiedocument

Op basis van het voorliggende document kunnen marktpartijen een reactie geven op de in deze marktconsultatie gestelde vragen. De Belastingdienst heeft hiertoe in dit document de behoefte omschreven en toegelicht volgens welke procedure deze marktconsultatie verloopt.

Indien u ondanks de zorg die de Belastingdienst aan dit document heeft besteed onduidelijkheden of onvolkomenheden aantreft of er bij u vragen rijzen, wilt u die dan bij de contactpersoon (zie paragraaf 2.2.) kenbaar maken.

De resultaten van de marktconsultatie kunnen worden verwerkt in aanbestedingsdocumenten en/of openbaar gemaakt, zonder dat deze herleidbaar zijn tot de deelnemers van de marktconsultatie. De Belastingdienst mag vrij over de verstrekte informatie beschikken en deze informatie gebruiken. De door u gegeven informatie zal vertrouwelijk worden behandeld.

2.4. Planning marktconsultatie

Activiteit	Datum
Verzenden marktconsultatie	Woensdag 17 mei 2023
Uiterst moment ontvangst beantwoording	12 juni, 17.00 uur
Dialoog Gesprekken	Nader te plannen

2.5. Vormvereisten

Het is uitdrukkelijk niet de bedoeling offertes in te leveren, wij verzoeken u dringend om u te beperken tot de beantwoording van de vraagstelling uit paragraaf 1.2. Daarnaast ontvangen wij graag van de contactpersoon aan uw kant de onderstaande gegevens:

Gegevens	
Organisatie	
Contactpersoon	
Functie	
Telefoonnummer	
E-mailadres	

2.6. Randvoorwaarden deelname aan deze marktconsultatie

Door deelname aan deze marktconsultatie gaat u onvoorwaardelijk akkoord met de alle voorwaarden zoals vermeld in dit marktconsultatiedocument. Hieronder staan de randvoorwaarden limitatief opgesomd:

- Deelnemen aan de marktconsultatie is voor marktpartijen én de Belastingdienst vrijblijvend;
- Dit marktconsultatiedocument is uitsluitend bestemd voor marktconsultatie doeleinden;
- De marktconsultatie is geen onderdeel van de mogelijke aanbesteding maar een consultatieronde;
- Door de deelnemers aan deze marktconsultatie kunnen geen kosten in rekening worden gebracht aan de Belastingdienst in verband met het uitbrengen van de reactie op deze marktconsultatie en de daarvoor uit te voeren werkzaamheden;
- Deze marktconsultatie dient uitdrukkelijk niet om een voorselectie te maken van marktpartijen in het kader van een mogelijke toekomstige aanbesteding. Marktpartijen die niet deelnemen aan de marktconsultatie zijn daarmee niet uitgesloten van deelname aan een mogelijk verwervingstraject. Evenmin zijn marktpartijen die deelnemen aan de marktconsultatie op enige wijze uitgesloten van, of bevoorrecht in, een eventuele deelname aan een aanbestedings- of verwervingstraject;
- De resultaten van de marktconsultatie worden mogelijk verwerkt in aanbestedingsdocumenten (bijvoorbeeld door tijdens deze marktconsultatie ontvangen informatie te verwerken in het Beschrijvend document), zonder dat deze herleidbaar zijn tot de deelnemers aan de marktconsultatie. Door deel te nemen aan marktconsultatie stemt u hiermee in;
- Verstrekte informatie en scope in het kader van de marktconsultatie kan (deels) afwijken van in de aanbestedingsprocedure te verstrekken informatie;
- De inbreng van deelnemers aan deze marktconsultatie zal zoveel mogelijk vertrouwelijk behandeld worden, waarbij in ieder geval rekening zal worden gehouden met de gerechtvaardigde belangen van de deelnemers van deze marktconsultatie en de Belastingdienst;
- De voertaal tijdens deze marktconsultatie is Nederlands, productinformatie mag in het Engels;
- De Belastingdienst is op geen enkele wijze gebonden aan de uitkomsten van de marktconsultatie;
- Deelnemers aan de marktconsultatie kunnen geen rechten ontleen aan deelname aan de marktconsultatie of aan enige informatie die tijdens de marktconsultatie wordt verstrekt;
- Het accepteren of honoreren van eventuele voorwaarden of claims (waaronder voorwaarden of claims t.a.v. Intellectuele Eigendomsrechten zoals auteursrecht) van deelnemers aan deze marktconsultatie over het gebruik van door hen aan de Belastingdienst in het kader van deze marktconsultatie informatie of vertrouwelijkheid en geheimhouding van door de betreffende marktpartij verstrekte informatie is niet mogelijk. Wel zal de Belastingdienst bij het verstrekken, openbaren of verwerken van informatie verkregen van deelnemers tijdens, of voortvloeiend uit, deze marktconsultatie zoveel mogelijk trachten te voorkomen dat deze informatie direct herleidbaar is tot een bepaalde deelnemer van de marktconsultatie.
- Indien intellectuele eigendomsrechten (waaronder auteursrechten) van een deelnemer of van derden van toepassing zijn dan wordt de Belastingdienst door de betreffende deelnemer aan deze marktconsultatie gevrijwaard van een vordering van deze deelnemer en/of derden op basis van die intellectuele eigendomsrechten (waaronder auteursrechten).
- De Belastingdienst behoudt zich het uitdrukkelijk het recht voor: De planning van deze marktconsultatie zoals opgenomen in dit document te wijzigen. Daarnaast ook om de marktconsultatie tijdelijk en/of definitief te staken.

Hoofdstuk 3. Context

3.1. Context

De Belastingdienst heeft de afgelopen jaren geïnvesteerd in een Analytics Capability. Een onmisbaar onderdeel van deze Capability is de beschikbaarheid van een centraal *Data Platform*. Het Data Platform biedt de functionaliteit om gegevens veilig op te slaan, op te werken en beschikbaar te stellen aan afnemende klantgroepen via speciaal ontwikkelde Analytics IT-services.

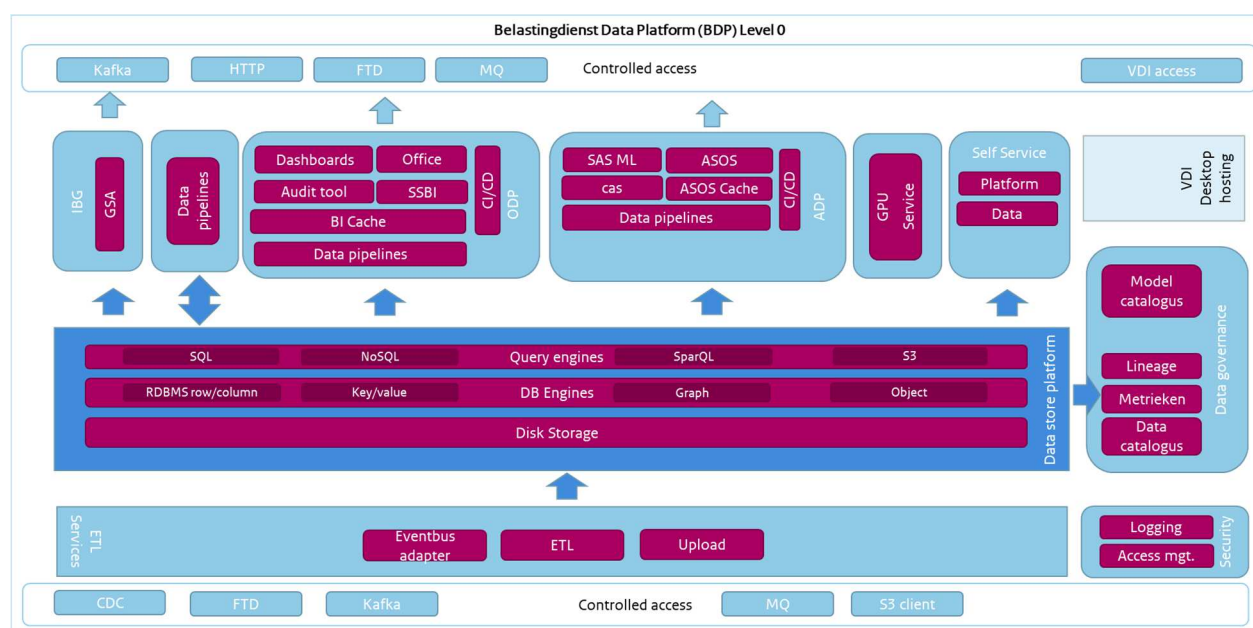
Op basis van deze Analytics IT-services zijn vele honderden informatieproducten ontwikkeld en beschikbaar gesteld aan afnemers binnen de Belastingdienst. Het huidige Data Platform wordt geleverd door de leverancier Teradata en bestaat o.a. uit on-premise hardware appliances, DBMS software en B&R apparatuur. Dit Data Platform wordt gebruikt door vele honderden gebruikers en wordt miljoenen keren per dag geraadpleegd door middel van query's. De volgende categorieën van informatieproducten zijn afhankelijk van het Data Platform en de Analytics IT-services:

- **Analytics informatieproducten:**
 - Beleidsondersteunende analyses;
 - Tactische- en Strategische MI rapportages en dashboards;
 - Risico- en signaalmodellen.
- **Niet-Analytics informatieproducten:**
 - Klantbeeld applicaties;
 - Gegevensleveringen aan derden;
 - Gegevens opwerking van bron t/m gegevens product (voor een informatieproduct);
 - Operationele Management Informatie en Operationele analyse.

Op dit moment worden er meer dan 200 administratieve systemen van de Belastingdienst structureel ontsloten. De beschikbaarheid van deze gegevens heeft ertoe geleid dat het Data Platform breder is ingezet dan alleen voor Analytics informatieproducten. Recent inzicht heeft ertoe geleid dat de behoefte aan data opslag niet alleen relationele opslag van gegevens betreft, maar ook opslag van ongestructureerde gegevens.

3.2. Beoogde Scope

De Belastingdienst is voornemens een nieuw **Data Storage Platform (DSP)** te verwerven dat het huidige data storage platform zal vervangen. De nieuwe oplossing zal ingezet worden als centrale gegevensinfrastructuur voor het opslaan, beheren en verwerken van gegevens die ondersteunend zijn aan de besluitvorming en informatievoorziening voor derden. De oplossing zal niet ingezet worden voor primaire systemen van de Belastingdienst. Onderstaand figuur beschrijft de scope (donkerblauw) van de uit te vragen oplossing in de context van de relevante gegevens en analytics ICT services.



Figuur 1: De scope van de uitvraag (donkerblauw) binnen het Belastingdienst Dataplatform (BDP).

3.3. Toelichting BDP

Bovenstaande figuur laat op hoofdlijnen de onderdelen van de doelarchitectuur zien. Het is één dataplatform voor alle use cases van secundair gebruik (met uitzondering van operationele monitoring). Secundair gebruik is gebruik dat ondersteunend is aan de primaire processen van de Belastingdienst. De onderdelen worden hieronder puntsgewijs toegelicht.

3.3.1. BDP

BDP staat voor Belastingdienst Data Platform. Het is een omgeving die voorziet in veilige opslag en gebruik van data. Alle ingaande en uitgaande toegang verloopt via gecontroleerde routes (controlled access). Hiervoor wordt gebruik gemaakt van technische voorzieningen die in het kader van de ADP al zijn ingericht (netwerkvirtualisatie en segmentering). De volgende koppelvlakken zijn gedefinieerd voor de BDP:

Afkorting:	Toelichting	Gebruikt voor ingaande dataroute?	Gebruikt voor uitgaande dataroute?
CDC	Change Data Capture voorziet in een 1-op-1 kopie van een bron. Dit is in eerste instantie de belangrijkste route om data in de datakuis van de SED te krijgen.	Ja	
FTD	File Transfer Dienst kan gebruikt worden om complete datasets af te leveren. Dit kan bijvoorbeeld een ontvangen bestand van derden zijn of een (delta)dump van een bron die niet met CDC te ontsluiten is. FTD kan gebruikt worden om complete datasets af te leveren richting een afnemend systeem (FTD is een bestaande file transfer dienst van de Belastingdienst);	Ja	Ja
KAFKA	Kafka is bedoeld om (near) real time events te leveren en te publiceren (bijvoorbeeld signalen);	Ja	Ja
MQ	MQ is alternatief voor bronnen die niet via Kafka kunnen aansluiten. Om berichten te verzenden naar een afnemend systeem	Ja	Ja
S3	Dit is een client waarmee bestanden in een bucket van een object store kunnen worden gezet. Bijvoorbeeld sets van documenten of een verzameling bestanden in het kader van een boekenonderzoek.	Ja	
HTTP	Koppelvlak voor webservices (pull) en interactief raadplegen (bijvoorbeeld een dashboard)		Ja
VDI	Om ontwikkelaars en beheerders toegang te geven tot werkgebieden & productgebieden wordt gebruik gemaakt van VDI's.	Ja	Ja

3.3.2. Data Storage Platform

Het data storage platform is de kern van het data platform. Deze voorziet in alle typen opslag die nodig zijn om de verschillende use cases te bedienen (zowel analytisch als operationeel). Via interfaces gebaseerd op open standaarden worden datasets uitgeleverd. Belangrijk is ook het koppelvlak naar Data Governance om operationele metadata en lineage vast te leggen. Verder is ook het raakvlak naar de data pipelines van belang. Hiermee worden de transformaties op het dataplatform aangestuurd. We beogen hiervoor gebruik te maken van in-database transformaties die via pipelines georchestreerd worden.

Het data storage platform voorziet in de volgende opslagmethoden:

- Relationale opslag (RDBMS) in row of column store. Voor tabelgeëïntendeerde data. Ondermeer de datakuis van de SED wordt hiermee ondersteund.
- Key/value store: Voor snel opslaan en teruglezen van data met een niet uniforme structuur. Bijvoorbeeld om events op te slaan of social media feeds.
- Graph store: In een graph store kan data worden opgeslagen die gerepresenteerd is als en graaf. Denk hierbij bijvoorbeeld aan een netwerk met personen en hun onderlinge relaties.
- Object store: Opslag om data als (collecties van) objecten op te slaan, met hun metadata. Bijvoorbeeld een set documenten of een collectie van audit files in het kader van een onderzoek. Het is eenvoudig schaalbaar en relatief goedkoop.



Belangrijke keuze is dat het data storage platform vooral bedoeld is voor het opslaan van data en het snel kunnen uitleveren van sets van data. In tegenstelling tot de huidige situatie zal het nieuwe data storage platform run-time niet gebruikt worden door de afnemende toepassingen. De type toepassingen zijn in bovenstaande afbeelding aan de bovenzijde getekend en worden uitgevoerd op een eigen applicatie hosting platform met een eigen data cache.

Verder wordt gekeken naar kostenefficiëntie. Data wordt opgeslagen op de economisch meest voordelige wijze. In de huidige situatie worden bijvoorbeeld nog reserve(kopieën) van data op de DWA opgeslagen. In de nieuwe situatie gebeurt dat in een object store. Dat is een beheerde omgeving waar collecties van bestanden geplaatst kunnen worden.

3.3.3. Toepassingen ADP, ODP, IBG

Bovenin de afbeelding zijn de typen toepassingsgebieden getekend. Toepassingen worden aangeboden op een eigen gecompartmenteerd virtueel hosting platform. Zij ontvangen via standaard koppelvlakken datasets vanaf het dataplatform.

- De ADP bevat gecompartmenteerde werkgebieden & productstraten voor analytisch gebruik. Deze kunnen gebruik maken van GPU acceleratie;
- De ODP volgt hetzelfde principe als de ADP, maar is bedoeld voor meer operationele analyses en verwerkingen. Ook daar kunnen tijdelijke werkgebieden en permanente productstraten worden ingericht.
- Data wordt beschikbaar gesteld via een aanvraagproces en kan uitsluitend via gecontroleerde dataroutes naar buiten worden gebracht. De aangeboden producten tools zullen echter verschillen van de ADP.
- Het inwinnen en beschikbaar stellen van gegevens (IBG) is ook een type toepassing op dit platform. Deze ontvangt zijn sets van de SED en levert deze uit via een gecontroleerde route richting de Poort. (bijvoorbeeld FTD). Omgekeerd kan IBG datasets vanuit Poort ontvangen voor verwerking in de SED.

3.3.4. Data Governance

Data Governance voorziet in inzicht in het bestaan en gebruik van datasets op het dataplatform. Belangrijke onderdelen zijn een administratie van operationele metadata en lineage informatie. Voor operationele metadata en lineage is momenteel [Datahub](#) beoogd.

Datahub is een open source oplossing wordt nu al gebruikt voor de setgebaseerde dienst. Dit product voorziet in het automatische inlezen van metadata en lineage informatie en voorzien in column leven lineage en visualisatie. Omdat metadata zelf ook weer data is en de producten open zijn, kunnen we deze metadata en lineageinformatie ook weer gebruiken in analyses om patronen in datagebruik en mogelijke kwaliteitsverbeteringen te detecteren. Deze informatie kan

via een push of een pull verkregen worden. Voor de pull optie is het nodig dat de DSP oplossing logt in een leesbaar formaat, dat ontsloten kan worden naar de voorziening voor operationele metadata.

3.3.5. ETL & Data Pipelines

De datatransformaties worden uitgevoerd met behulp van georchestreeerde data pipelines. Deze data pipelines worden in-database uitgevoerd. De data pipelines worden zoveel mogelijk modelgedreven uitgegenereerd (op basis van datamodellen en open standaarden). De transformaties zelf vinden in de databases plaats. Dit stelt een eis aan de uit te vragen oplossing voor de DSP, namelijk dat deze ingericht moet zijn om grootschalig in-database transformaties op sets van data te ondersteunen. Ook voor de ETL service (die feitelijk ELT doet) willen we model gestuurde data pipelines inzetten.

3.3.6. Self service

Het beschikbaar stellen van data en toepassingen gebeurt zo veel mogelijk op basis van een self service portaal, net als dit nu voor de ADP gebeurt. Dit betekent dat omgevingen worden aangevraagd in een management portaal. Data wordt aangevraagd via het data dashboard. De aansturing van de uitrol doen we met playbooks op ansible. De oplossing voor DSP moet over koppelvlakken beschikken om deze automation in te kunnen richten. Denk daarbij bijvoorbeeld aan het aanmaken van databases, schema's en autorisaties.

Security

Het controleren van toegang tot data en loggen van gebruik is nu ingericht op de ADP. Dit is echter een belangrijk aspect van alle gebruik van data. Het nieuwe dataplatform zal voor alle vormen van gebruik autorisatie en logging inrichten. Dit betekent dat het DSP logininformatie moet kunnen pushen, of loggen in een leesbaar formaat, zodat log-agents deze informatie kunnen doorsturen naar de centrale logvoorziening (Splunk).

3.4. Gehanteerde architectuur Principes

De Belastingdienst is voornemens een nieuw **Data Storage Platform (DSP)** te verwerven dat het huidige data storage platform zal vervangen. De nieuwe oplossing zal ingezet worden als centrale gegevensinfrastructuur voor het opslaan, beheren en verwerken van gegevens. Daarmee is het DSP het hart van het Belastingdienst Data Platform (BDP).

#	Uitgangspunt	Rationale	Gevolgen voor de gevraagde oplossing
1	Componenten (van de oplossing) zijn tegen aanvaardbare kosten vervangbaar voor andere functioneel gelijkwaardige componenten.	Technologie ontwikkelt zich voortdurend. Hard- en software componenten moeten na verloop van tijd vervangen worden. Hetzij vanwege life cycle management, hetzij vanwege aanbestedingsregels. Deze veranderingen moeten niet leiden tot een grote migratie inspanning.	- Scheiding van verantwoordelijkheden is belangrijk binnen het data storage platform. Een component wordt alleen ingezet waarvoor het is bedoeld. Alleen dan kan deze vervangen worden zonder dat onbedoeld functionaliteit verdwijnt. - Grote ontwikkelinspanning in proprietary oplossingen willen we voorkomen. Er moet expliciet rekening gehouden worden met migreerbaarheid in de toekomst. Als proprietary code wordt gebruikt dan alleen als het niet anders kan én deze gegenereerd wordt.

2	Toekomstige functionaliteit is inpasbaar	De wereld rondom data analyse blijft sterk ontwikkeling. Komende jaren zullen er nieuwe toepassingen komen die ook voor de Belastingdienst interessant zijn. Deze toepassingen willen we eenvoudig kunnen introduceren naast bestaande toepassingen waarbij zij aansluiten op reeds bestaande koppelvlakken. Ook willen we oudere componenten kunnen uitfasen terwijl nieuwere blijven draaien.	- Scheiding van verantwoordelijkheden is belangrijk binnen het dataplatform. Functionele componenten communiceren met de dataopslag via vastgestelde interface via open standaarden. Functionaliteit wordt op de juiste plaats belegd. - Functionele componenten hebben geen koppelvlakken naar elkaar (tenzij expliciet ontworpen). Hiermee zouden immers nieuwere componenten functioneel afhankelijk kunnen worden van oudere componenten die uitgefaseerd moeten worden
3	We maken gebruik van open standaarden en open source, tenzij (uitgangspunt rijk)	Er is veel code ontwikkeld die te maken heeft met opwerken en transformeren van data. Dit leidt tot een technische lock-in omdat dit niet eenvoudig om te zetten is naar een ander platform. Gebruik van open standaarden verkleint de kans op een lock-in	- De requirements voor de aanbesteding van het Data Storage Platform zullen op de koppelvlakken open standaarden voorschrijven. - Oplossingen die gebaseerd zijn op open source en open standaarden worden mogelijk extra punten aan toegekend; - Voor de ontwikkeling van Data Pipelines wordt gebruik gemaakt van open standaarden en open source.
4	Data wordt opgeslagen op de economisch meest voordelige wijze	Veel data bestaat uit kopieën van afgeleide bronnen. Het is niet zo dat alle gebruikers, altijd met alle data analytics aan het bedrijven zijn. Het is daarom niet doelmatig om al die data op een analytical appliance te plaatsen.	- Er komt één data storage platform waarin we alle vormen van data (gestructureerd en ongestructureerd) kunnen opslaan. Dit platform gaat ook de setgebaseerde dienst ondersteunen. Dit platform is geoptimaliseerd voor het snel leveren van datasets. Het bevat geen analytical accelerators. - Optimalisatie vindt plaats per type gebruik. Voor analytics middels een analytical cache waarin datasets klaargezet worden voor gebruik. Voor operationeel gebruik wordt per use case /toepassing geoptimaliseerd.
5	Data is een asset waarvan de kwaliteit continue bestuurd moet worden. Automatisering is randvoorwaarde voor succes.	De belastingdienst gebruikt de data op het dataplatform voor beslisondersteuning en diverse operationele doelen. Het is belangrijk dat kwaliteit van data past bij het gebruiksdoel. Hiervoor is het nodig dat er zicht is op de aanwezige data-assets, hun wijze van ontstaan en de kwaliteit. Gezien de omvang van de datasets is handmatig bijhouden hiervan niet doenlijk.	- Er moet een haalbaar minimaal niveau van datakwaliteit gedefinieerd worden. Het nieuwe platform bevat tools die klantgroepen helpen om continue, incrementeel naar hogere datakwaliteit toe te werken. “Observability “ en automatisering van operationele metadata is een belangrijke functie voor het nieuwe platform

Bijlage 1: Begrippenlijst

Begrip	Definitie	Concrete voorbeelden
Access Management	Het beheer van toegang tot gegevens en bronnen, zoals gebruikersaccounts, wachtwoorden en autorisaties.	Active Directory, OAuth, Amazon IAM
Active Directory Integration	De integratie van Active Directory (een directory service van Microsoft) in andere toepassingen en systemen, waardoor gebruikers- en toegangsbeheer kan worden gecentraliseerd en geautomatiseerd.	AWS AD Connector, Azure Active Directory, Okta
API interface	Een interface die softwaretoepassingen met elkaar verbindt, waardoor gegevens en functionaliteit kunnen worden uitgewisseld.	REST API, GraphQL, SOAP
Block-level interface	Een interface waarmee applicaties toegang hebben tot individuele blokken van gegevens op een opslagapparaat, waardoor gegevens kunnen worden gelezen en geschreven.	iSCSI, Fibre Channel, Amazon EBS
Blocklevel Storage	Een vorm van opslag waarbij gegevens worden opgeslagen in blokken en toegankelijk zijn via een blokprotocol.	iSCSI, Fibre Channel, Amazon EBS
Column Storage	Een vorm van opslag waarbij gegevens worden opgeslagen in kolommen in plaats van rijen, waardoor efficiëntere query's mogelijk zijn voor bepaalde werklasten.	Apache Cassandra, Apache HBase, Amazon Redshift
Data Deduplication	Een technologie die redundante gegevens verwijdert om opslagruimte te besparen, zonder verlies van gegevensintegriteit of -beschikbaarheid.	NetApp deduplication, EMC Data Domain, IBM Spectrum Protect
Data Encryption (at rest)	Het proces van het coderen van gegevens om vertrouwelijkheid en beveiliging te bieden, vaak via algoritmen zoals AES en RSA.	OpenSSL, BitLocker, AWS KMS
Data Masking	Een technologie die gevoelige gegevens vervangt door irrelevante of fictieve gegevens om de privacy en veiligheid van gegevens te beschermen, terwijl de bruikbaarheid wordt behouden.	Delphix, Informatica, IBM Guardium
Data Retention	Het beleid en de praktijk van het bewaren van gegevens voor een bepaalde periode voor compliance, juridische of bedrijfsdoeleinden.	GDPR, HIPAA, PCI DSS
Physical Storage Hardware	Fysieke apparaten die worden gebruikt om gegevens op te slaan, zoals harde schijven, solid-state drives en tapes.	Western Digital, Seagate, Toshiba
Data Storage Platform	Een geïntegreerde oplossing voor het opslaan, beheren en beschermen van gegevens.	Amazon S3, Microsoft Azure, Google Cloud
Data Storage Software	Software die wordt gebruikt om gegevens op te slaan, te beheren en te beschermen op opslagapparaten.	NetApp ONTAP, Dell EMC Unity, IBM Spectrum Storage Suite
Data Virtualization	Een technologie die gegevens van meerdere bronnen samenbrengt en deze presenteert als één virtuele gegevensbron, waardoor het gemakkelijker wordt om gegevens te analyseren en te gebruiken.	Denodo, Informatica, AWS Glue
Event Storage	Een vorm van opslag die is ontworpen om gebeurtenisgegevens vast te leggen, zoals logbestanden en IoT-gegevens.	Splunk, Elastic Stack, IBM Log Analysis
File Storage	Een vorm van opslag die is ontworpen voor het opslaan en ophalen van bestanden, vaak via netwerkshares.	NFS, SMB, Amazon EFS
Graph interface	Een interface waarmee gebruikers Graph-query's kunnen uitvoeren op Graph databases.	Neo4j, ArangoDB, Amazon Neptune

Graph Storage	Een vorm van opslag die is ontworpen om relaties tussen gegevenspunten te modelleren, zoals sociale netwerken en aanbevelingssystemen.	Neo4j, Amazon Neptune, Microsoft Azure Cosmos DB
Key-value Storage	Een vorm van opslag waarbij gegevens worden opgeslagen als sleutel-waarde paren, waardoor snelle opzoeking mogelijk is.	Redis, Riak, Amazon DynamoDB
Logging	Het bijhouden van activiteiten en gebeurtenissen in een systeem of toepassing, zoals fouten, waarschuwingen en beveiligingsgebeurtenissen.	Splunk, ELK Stack, AWS CloudTrail
Native interface	Een interface die specifiek is ontworpen voor een bepaalde database of opslagengine, waardoor optimale prestaties en efficiëntie worden bereikt.	MongoDB native driver, Redis native client, Cassandra native driver
Object Storage	Een vorm van opslag waarbij gegevens worden opgeslagen als objecten met metadata, waardoor schaalbaarheid en flexibiliteit mogelijk zijn.	Amazon S3, Microsoft Azure Blob Storage, Google Cloud Storage
Relational Storage	Een vorm van opslag waarbij gegevens worden opgeslagen in tabelvorm en worden gerelateerd aan elkaar door middel van sleutels.	MySQL, Oracle Database, Microsoft SQL Server
S3 interface	Een interface waarmee gebruikers objecten kunnen opslaan, ophalen en beheren in Amazon S3 object storage.	AWS SDK, Cyberduck, S3 Browser
Scalable Hardware	Hardware die is ontworpen om te kunnen worden geschaald om te voldoen aan de groeiende eisen van gegevensopslag en -verwerking, vaak met behulp van technologieën zoals clustering en virtualisatie.	Cisco UCS, Dell EMC VxRail, NetApp FlexPod
SparQL interface	Een interface waarmee gebruikers SparQL-query's kunnen uitvoeren op RDF Graph databases.	Virtuoso, Stardog, Apache Jena
SQL interface	Een interface waarmee gebruikers SQL-query's kunnen uitvoeren op relationele databases.	MySQL, Oracle Database, Microsoft SQL Server
Structured Data	Gegevens die worden opgeslagen in een vooraf gedefinieerde structuur, zoals tabellen in een relationele database.	Klantgegevens in een CRM-systeem
Structured Data Access	De manier waarop gestructureerde gegevens worden opgehaald en verwerkt, vaak via SQL-query's.	SELECT * FROM klanten WHERE land='Nederland'
Structured Storage Management	Het beheer van opslag van gestructureerde gegevens, zoals het maken van back-ups, het herstellen van gegevens en het beheren van opslagruimte.	NetApp OnCommand, Dell EMC Unity, IBM Spectrum Control
Tiered Storage	Een strategie voor het optimaliseren van de kosten en prestaties van opslag door gegevens te verplaatsen naar verschillende opslag-tiers op basis van hun leeftijd, waarde en toegankelijkheidseisen.	HSM, NetApp FabricPool, Dell EMC CloudArray
Unstructured Data	Gegevens die geen vaste structuur hebben, zoals tekst, afbeeldingen, audio en video.	Sociale mediaberichten, e-mails, videobestanden
Unstructured Data Access	De manier waarop ongestructureerde gegevens worden opgehaald en verwerkt, vaak met behulp van zoekopdrachten.	zoeken naar trefwoorden in een e-mailarchief
Unstructured Storage Management	Het beheer van opslag van ongestructureerde gegevens, zoals het maken van back-ups, het herstellen van gegevens en het beheren van opslagruimte.	Dell EMC Isilon, NetApp StorageGRID, IBM Cloud Object Storage