



IMPLANTACIÓ PLATAFORMA DENODO PER A LA VIRTUALITZACIÓ I GOVERN DE DADES ANALÍTIQUES

Expedient número: 15012664

PLEC DE CONDICIONS TÈCNIQUES

Setembre 2023

Carles Teixidó Gálvez

RESPONSABLE PROJECTES PRESTACIÓ SERVEI

ÍNDEX

1. INTRODUCCIÓ I OBJECTIU	4
2. REQUISITS FUNCIONALS.....	7
2.1. Connectivitat - Fonts de Dades	7
2.1.1. Requisits generals.....	7
2.1.2. Detall de Requisits	7
2.2. Modelat de Dades	8
2.2.1. Requisits generals.....	8
2.2.2. Detall de Requisits	8
2.3. Lliurament de dades.....	10
2.3.1. Requisits generals.....	10
2.3.2. Detall de Requisits	10
2.4. Catàleg de Dades	11
2.4.1. Requisits generals.....	11
2.4.2. Detall de Requisits	11
2.5. Seguretat/Governança de Dades	12
2.5.1. Requisits generals.....	12
2.5.2. Detall de Requisits	13
2.6. Optimització i Rendiment.....	14
2.6.1. Requisits generals.....	14
2.6.2. Detall de Requisits	14
2.7. Memòria cau / Replicació Selectiva	15
2.7.1. Requisits generals.....	15
2.7.2. Detall de Requisits	16
2.8. BigData / Cloud.....	16
2.8.1. Requisits generals.....	17
2.8.2. Detall de Requisits	17
2.9. Semàntica i Governança de Dades	17
2.8.1. Requisits generals.....	17
2.8.2. Detall de Requisits	17
2.10. Processament per lots.....	18
2.10.1. Requisits generals.....	18

2.10.2.	Detall de Requisits.....	18
2.11.	Desenvolupament i Operacions (DevOps)	19
2.11.1.	Requisits generals.....	19
2.11.2.	Detall de Requisits.....	19
2.12.	Desplegaments.....	21
2.12.1.	Requisits generals.....	21
2.12.2.	Detall de Requisits.....	21
3.	REQUISITS NO FUNCIONALS	22
4.	SERVEIS A CONTRACTAR	23
4.1.	Subscripció.....	23
4.2.	Servei d'Instal·lació.....	23
4.3.	Servei de desenvolupament.....	24
5.	SERVEI	25
5.1.	RESPONSABILITAT DE L'OFERTANT	25
5.2.	LLOC DE REALITZACIÓ DELS TREBALLS	25
5.3.	SEGUIMENT DEL SERVEI	25
5.4.	IDIOMA	25
5.5.	GARANTIA.....	25
5.6.	SEGURETAT I CONFIDENCIALITAT	26
5.6.1.	Confidencialitat i publicitat del servei	26
5.6.2.	Propietat intel·lectual	26
5.6.3.	Seguretat i protecció de dades	27

1. INTRODUCCIÓ I OBJECTIU

Avui dia, les arquitectures centralitzades no són suficients per respondre a les canviants necessitats de negoci, les regulacions de privadesa o la vertiginosa evolució de la tecnologia, entre d'altres.

La clau per donar resposta a totes les necessitats d'informació de les organitzacions avui dia és a les arquitectures distribuïdes i lògiques. La virtualització de dades és la base d'aquestes arquitectures:

- **Distribuïdes:** avui dia, els volums de dades són massa grans i aquests estan massa distribuïts. No és factible (ni desitjable) tenir totes les dades en un sol sistema. La realitat és que les dades residiran en múltiples sistemes/ubicacions. A més, les necessitats analítiques i operacionals modernes són massa diverses, per la qual cosa no és realista pensar en un sol sistema (“one size never fits all”).
- **Lògiques:** els consumidors de dades es beneficien en accedir a les dades a través de models semàntics, desvinculats de la ubicació real de les dades i els esquemes de dades físiques. Mitjançant aquest enfocament semàntic, els models de dades s'adapten a les necessitats dels consumidors i no a l'inrevés. A més, a través d'un model semàntic és possible aplicar polítiques comunes sobre fonts de dades disperses, que altrament no és possible a un cost i esforç raonables, si s'han d'aplicar a totes i cadascuna de les fonts. Finalment, aquest desacoblament proporciona independència i autonomia als propietaris de la informació, que poden evolucionar la tecnologia i fer canvis en la seva infraestructura sense problemes, per exemple, iniciar una transició al núvol.

L'objectiu a assolir amb la introducció d'una plataforma de virtualització de dades és desacoblar l'heterogeneïtat de fonts de dades de les aplicacions que consumeixen, proporcionant un accés homogeni, unificat i governat a totes dades de l'organització, internes o externes, a qualsevol ubicació, per satisfer totes les necessitats de negoci, tant operacionals com informacionals.

La virtualització de dades haurà de proporcionar a les aplicacions consumeixen un únic punt lògic per accedir a tot tipus de dades, evitant connexions punt a punt complexes i difícils de mantenir, i desacoblant els consumidors de dades de les seves fonts. La capa de virtualització de dades ha de permetre generar múltiples vistes de negoci, lògiques, sobre les mateixes dades físiques per adaptar-se als requisits específics de cada usuari i/o aplicació de negoci, és a dir, amb convencions d'anomenat, format i esquema específiques i lliurant les dades en el format més adequat per a cadascun, per exemple, SQL o REST/JSON. La solució haurà de proporcionar simultàniament múltiples visions sense duplicar les dades ni copiar-les obligatòriament a un repositori centralitzat.

Per tant, la plataforma haurà de ser la peça clau que permeti desplegar arquitectures lògiques i distribuïdes, que donin resposta als reptes presents i futurs a què s'enfronten les organitzacions pel que fa a la gestió de les dades.

Com a capa de desacoblament lògica, la virtualització de dades haurà de facilitar la sostenibilitat dels consumidors de dades, apantallant-los davant de canvis que es produeixin als orígens de dades (per exemple, per l'actualització d'un sistema).

La plataforma virtualització de dades haurà de proporcionar les següents característiques clau:

- A. **Abstracció:** les aplicacions de negoci han d'accedir a les dades mitjançant un únic punt lògic d'accés. Aquest punt amagarà la complexitat de les diferents tecnologies i models d'informació subjacents.
- B. **Capa semàntica unificada:** l'usuari haurà de poder definir les entitats de negoci en aquesta capa representant una definició comuna per a tota l'empresa, i exposant-les com a serveis de dades que qualsevol aplicació pot consumir evitant inconsistències i millorant la qualitat general de la informació.
- C. **Federació de dades:** la plataforma haurà de permetre combinar dades de diferents orígens, federant-hi l'accés a través de la capa de virtualització de dades, sense copiar-les en un repositori centralitzat. L'usuari evita així replicar dades d'un origen a un altre.
- D. **Portar el processament a les dades:** la plataforma haurà d'aprofitar al màxim les orígens de dades especialitzades, explotant les seves capacitats de processament (per exemple, bases de dades analítiques que poden executar agregacions sobre grans volums de dades de manera eficient, motors de llacs de dades que poden executar consultes sobre col·leccions de fitxers emmagatzemats al núvol, etc.). La capa de virtualització actuarà com un node mestre que intel·ligentment distribueix l'execució de la consulta entre els diferents orígens fent push-down de les consultes i minimitzant el trànsit de dades a través de la xarxa.
- E. **Gestió de canvis:** els canvis a les fonts de dades oa les tecnologies subjacents es podran acomodar fàcilment en aquesta capa, minimitzant l'impacte sobre les aplicacions consumidores. L'organització, o diferents parts de la mateixa, o entitats externes, tindran total independència per evolucionar els seus sistemes amb nous components tecnològics (cosa molt freqüent al món del Big Data).
- F. **Capa de seguretat i governança unificada:** la plataforma haurà de poder definir, administrar i fer complir regles de seguretat i governança de dades des d'un sol punt, i definir-les en termes del model canònic d'entitats de negoci comunes per a tota l'organització.
- G. **Facilitat d'ús:** la solució haurà de ser fàcil d'usar. El procés de connectar-se a fonts de dades, combinar-les per crear entitats de negoci i publicar-les a través de múltiples protocols, s'ha de poder fer de manera gràfica.
- H. **Flexibilitat:** la plataforma ha de ser flexible per donar cabuda a nous models i adaptar-se de manera àgil a les necessitats de negoci.

- I. **Rendiment:** la plataforma haurà de proporcionar rendiment òptim mitjançant la combinació de tècniques d'optimització avançades (push-down), possibilitat d'ús de memòria cau i acceleració de consultes. La plataforma haurà de tenir un motor d'optimització basat en costos per seleccionar automàticament, en temps d'execució, la millor estratègia entre totes les possibles sense dependre de configuració ad-hoc.
- J. **Característiques enterprise:** capacitat de proporcionar de manera àgil i senzilla a cada consumidor de dades de l'organització la informació que necessita, en els formats que necessiten, garantint alhora la coherència, la governança, la qualitat i la seguretat de les dades.

La Plataforma Denodo dona resposta a totes aquestes necessitats que es detallen a continuació i que són indispensables per a la correcta integració a l'ecosistema de productes de TMB.

2. REQUISITS FUNCIONALS

A continuació, es descriuen amb més detall les funcionalitats que ha de proporcionar la solució:

2.1. Connectivitat - Fonts de Dades

2.1.1. Requisits generals

1. Heu de proporcionar una capa d'abstracció completa de tots els repositoris de dades de l'empresa, creant un únic punt d'accés des del qual accedir a qualsevol conjunt de dades, sense necessitat de copiar o replicar cap dada.
2. Heu de disposar d'una àmplia gamma de fonts, formats i tipus de dades suportades. Heu de permetre la possibilitat d'ampliar la llista de connectors de fàbrica amb connectors personalitzats.
3. Els connectors han de maximitzar la delegació per aprofitar al màxim les capacitats de processament de les fonts de dades.

2.1.2. Detall de Requisits

1. Bases de dades relacionals, Obligatòries: Oracle, SQLServer i MySQL. Tant "on premise" com al núvol.
2. Data warehouses i sistemes multidimensionals, Obligatòries: SAP BW i MS Analysis Services
3. Bases de dades en memòria, Obligatòria: SAP HANA, Oracle in-memory
4. Llacs de dades (data lakes) i Lakehouses, incloent Hive, SparkSQL, Obligatòria: Impala, Kudu, Presto
5. Data warehouses, data lakes i lakehouses al núvol, incloent Redshift, Snowflake, AWS RDS, Databricks Delta, Azure Synapsis, Obligatòria: Google BigQuery
6. Fitxers delimitats, logs, Excel, SAS
7. Fitxers XML, JSON
8. Fitxers en magatzems d'objectes distribuïts, incloent-hi HDFS, AWS S3, Azure ADLS, Google Cloud Storage
9. APIs SaaS, incloent SFDC, Google Analytics, ServiceNow, Dynamics 365, Marketo, Twitter

10. APIs Web Services (SOAP, REST, OData) amb formats de representació XML/JSON
11. Creació de fonts de dades REST a partir d'una especificació OpenAPI/Swagger
12. Aplicacions paquetitzades, incloent SAP, Oracle E-Business
13. Bases de dades NoSQL databases incloent, MongoDB, Cassandra, Neo4j
14. Integració amb LDAP / Active Directory
15. Fonts Streaming/IoT, Obligatoria Kafka, JMS
16. Extensibilitat: SDK per a desenvolupament de connectors ad-hoc
17. Suport de lectura/escriptura a les fonts

2.2. Modelat de Dades

2.2.1. Requisits generals

1. Ha de permetre la creació de vistes integrades que combinin dades de qualsevol font/tipus de dades, ja sigui al núvol, a les instal·lacions del client o distribuïts en diferents àrees geogràfiques.
2. Ha de proporcionar una eina de modelatge gràfica, amb gran facilitat d'ús i zero codi.

2.2.2. Detall de Requisits

1. El producte ha de proporcionar una eina de modelatge que inclou tots els passos de la creació d'un model de dades lògiques (és a dir, des de la connexió a les fonts de dades, la combinació i transformació per crear vistes lògiques i finalment la publicació de les mateixes per accedir-hi en múltiples formats)
2. El producte ha de proporcionar un enfocament de modelatge relacional utilitzant assistents gràfics per a operacions comunes de SQL (unions, unions, agregacions, projeccions, aplanament, etc.)
3. El producte ha de poder integrar en un sol model diferents tipus de fonts (per exemple, estructurades, semiestructurades, API) de diferents ubicacions (per exemple, a les instal·lacions, al núvol)
4. Ha de suportar els següents tipus de dades: suport per a tipus de dades estàndard (sencers, cadena de text, dates, etc.) i tipus de dades complexes (registres, arrays). S'han de poder modelar estructures de dades jeràrquiques comunament utilitzades a les API

5. Perfilat de dades: L'eina ha de poder importar índexs, PK, FK, estadístiques (informació bàsica de perfilat), descripcions, etc. de les fonts de dades
6. S'han de poder mapejar les capacitats de consulta d'origen. El model representa les capacitats de consulta de les fonts de dades i les té en compte en l'execució de les consultes
7. S'han de poder modelar les relacions entre fonts
8. S'han de poder crear modelatges automàtics basats en les metadades.
9. Heu de suportar múltiples models virtuals. Els models virtuals hauran de poder fer servir vistes de qualsevol altre model virtual
10. Hi ha d'haver una biblioteca de funcions de transformació: manipulació de cadenes de text, manipulació de dates, matemàtiques, trigonòmètriques, expressions regulars, xml/json, etc.
11. S'han de poder estendre les funcions personalitzades (UDF) i transformacions
12. S'ha de poder executar codi ja creat a la font de dades (UDF/SP). El producte ha de suportar l'execució de codi nadiu a les bases de dades a què està connectat. Per exemple, per aprofitar consultes optimitzades, funcions definides per l'usuari o procediments emmagatzemats ja disponibles.
13. S'ha de poder etiquetar el model (vistes i columnes) per definir polítiques d'accés que s'hi basen.
14. El producte ha de proporcionar una alternativa a l'eina de modelatge gràfica, apta per a usuaris avançats, per exemple, a través de seqüències d'ordres.
15. Ha de permetre la captura automàtica de meta-informació a les fonts de dades connectades
16. El producte s'ha d'integrar amb eines de modelatge de tercers, incloent-hi ER/Studio, ERwin, PowerDesigner
17. Cal poder versionar els models. L'eina ha de suportar diferents versions d'un model alhora
18. Heu de suportar les funcions geoespacial
19. Ha de tenir suport per a procediments emmagatzemats
20. Ha de tenir suport complet a CRUD (Crear, Llegir, Actualitzar, Eliminar)
21. Ha de suportar les transaccions distribuïdes

2.3. Lliurament de dades

2.3.1. Requisits generals

1. Heu de disposar de múltiples opcions de lliurament de dades per satisfer les necessitats del client a través d'interfícies SQL (JDBC, ODBC i ADO.Net), MDX per a anàlisis OLAP, API.
2. Heu de proporcionar una capa completa de serveis de dades amb suport OAuth 2.0, SAML, OpenAPI, OData 4, GraphQL, per a una fàcil interoperabilitat amb els sistemes en núvol.
3. Ha de disposar de la capacitat per desenvolupar i desplegar API de dades (DaaS) en minuts, sense codi.

2.3.2. Detall de Requisits

1. Cal permetre el lliurament de dades en format SQL a través de protocols JDBC, ODBC i ADO.NET
2. S'ha de permetre el lliurament de dades en forma d'API/serveis web mitjançant protocols SOAP, RESTful, OData
3. Cal permetre el lliurament de dades en format GraphQL.
4. Documentació de serveis de dades. El producte ha de proporcionar documentació OpenAPI (Swagger) dels serveis de dades publicats.
5. Cal permetre el lliurament de dades asíncrona/basada en esdeveniments a través de cues de missatges JMS i Kafka
6. S'ha de permetre l'accés directe des d'informes i/o eines de BI, incloent-hi Tableau, Obligatori: PowerBI, Spotfire, BO, OBIEE, Microstrategy
7. S'ha de permetre l'accés des d'eines de científics de dades, incloent-hi Notebooks
8. S'ha de suportar el moviment massiu de dades des de les fonts a un emmagatzematge diferent, per satisfer necessitats de migració de dades
9. S'han de poder crear exportadors personalitzats
10. S'ha de permetre el lliurament de dades de forma multidimensional a través de MDX. El producte ha de publicar dades a través del protocol MDX/XMLA per anàlisi OLAP

11. Cal permetre el lliurament de dades en format GeoJSON
12. Heu de suportar operacions d'IDU (inserció, esborrament, actualització). El producte ha de suportar operacions IDU a través de totes les interfícies disponibles (SQL, SOAP, REST)

2.4. Catàleg de Dades

2.4.1. Requisits generals

1. El producte ha de proporcionar una eina de catàleg de dades que ofereixi capacitats de descobriment de dades i autoservei per a usuaris de negoci, que els permeti explorar tant metadades com a dades en una única eina web.
2. S'ha de poder integrar amb catàlegs de dades de tercers i eines de visualització i informes.
3. El Catàleg de dades ha de proporcionar funcions de descoberta interactiva, capacitats de col·laboració i recomanacions personalitzades.

2.4.2. Detall de Requisits

1. El producte ha de proporcionar un catàleg complet de dades disponibles per al consum a tota l'organització
2. El catàleg de dades ha d'estar completament integrat amb la capa de lliurament de dades. El producte ha de garantir que el que veuen els usuaris al catàleg de dades és el que obtenen quan accedeixen a la capa de lliurament de dades
3. El producte ha de suportar relacions entre conjunts de dades i llinatge de dades
4. Semàntica. El producte ha de suportar la classificació dels elements (per exemple, en etiquetes o categories)
5. L'eina ha d'ajudar els usuaris a descobrir i explorar conjunts de dades i navegar a través de les relacions
6. Metadades: L'eina ha de suportar metadades actives (p. ex., estadístiques d'ús, perfilat de dades)
7. Metadades: S'han de poder ampliar les metadades dins del catàleg afegint propietats/camps personalitzats addicionals.

8. Metadades: Heu de permetre editar i enriquir la informació del catàleg per part d'usuaris autoritzats
9. Cerca: Heu de disposar de cerques avançades de metadades per trobar ràpidament conjunts de dades rellevants dins del catàleg.
10. Cerca: L'eina ha d'admetre la cerca a les dades a més de a les metadades (és a dir, al contingut en lloc de a les metadades)
11. Autoservei: La interfície de consulta ha de ser molt intuïtiva i fàcil d'usar per a usuaris no tècnics
12. Consum de dades: L'eina ha de permetre als usuaris obtenir una vista prèvia de les dades
13. Col·laboració i governança: L'eina ha de permetre als usuaris donar suport, advertir o alertar altres usuaris sobre conjunts específics de dades
14. Automatització basada en IA: L'eina ha de proporcionar funcions d'aprenentatge automàtic (ML) que analitzin l'activitat de l'usuari per oferir recomanacions personalitzades i dreceres als conjunts de dades més significatius
15. L'eina ha de proporcionar estadístiques d'ús de conjunts de dades, incloent-hi quins usuaris o aplicacions consulten quins conjunts de dades i quines consultes es llancen amb més freqüència
16. Integració: L'eina s'ha de poder integrar amb altres eines d'administració i de govern de dades de tercers, per exemple Colibra.
17. Consum de dades: L'eina ha de suportar l'exportació de dades seleccionades des del catàleg per ser analitzades localment o per eines externes, per exemple Excel, CSV, etc.
18. Personalització: La interfície d'usuari del catàleg s'ha de poder personalitzar per adaptar-la a l'organització

2.5. Seguretat/Governança de Dades

2.5.1. Requisits generals

1. Capa unificada de governança i seguretat. La plataforma ha de permetre definir, administrar i fer complir les regles de seguretat i el govern de dades des d'un únic punt. A més, les polítiques esmentades es defineixen en termes d'un model canònic d'entitats de negoci.

2. Heu de disposar d'un suport complet de governança i seguretat, incloent emmascarament de dades avançat i control d'accés basat en rols (RBAC). Disponible per a tots els actius de dades de l'organització, a través d'un únic punt de control i administració.

2.5.2. Detall de Requisits

1. Autenticació: Heu de disposar d'una gestió d'usuaris integrada
2. Autenticació: S'ha d'integrar amb LDAP / Active Directory
3. Autenticació: Ha de permetre autenticació avançada per a serveis web i APIs (p. ex., SAML, OAuth2, SPNEGO)
4. Autenticació: Heu de permetre l'inici de sessió únic (SSO) a tots els clients
5. Autenticació: Ha de permetre la transferència de credencials a fonts (pass-through) per aprofitar les polítiques de seguretat definides als orígens de dades
6. Autenticació: S'ha d'integrar amb password vaults externs, per exemple, CyberArk, per emmagatzemar les contrasenyes per accedir als orígens de dades
7. Autenticació: Heu de permetre emmagatzemar atributs de la sessió (per exemple, el departament de l'usuari o la ubicació actual) durant l'autenticació per utilitzar-los més tard per a l'autorització dinàmica basada en aquests atributs, o per inicialitzar els atributs de la connexió a les fonts de dades.
8. Autorització: Heu de permetre l'autorització basada en rols RBAC
9. Autorització: Ha de donar suport a l'herència de rols
10. Autorització: Heu de proporcionar un control d'accés a nivell de detall. Restriccions a nivell de columna i fila. Seguretat a nivell de cel·la.
11. Autorització: Ha de disposar de polítiques de seguretat coherents a tots els mecanismes de lliurament de dades. El producte permet la definició de polítiques de seguretat una vegada i les aplica a totes les interfícies (per exemple, SQL, API REST, catàleg de dades)
12. Privadesa: Ha de donar suport a l'emascament de dades. L'eina ha de permetre escollir com emascarar els "valors confidencials", per exemple, els primers/últims caràcters, establir els valors numèrics en 0 o 1, aplicar expressions personalitzades o establir valors nuls.
13. Privadesa: Ha de donar suport al xifratge de dades
14. Extensibilitat: Heu de proporcionar un SDK per definir polítiques d'autorització personalitzades

2.6. Optimització i Rendiment

2.6.1. Requisits generals

1. L'eina ha de disposar de tècniques d'optimització de consultes de darrera generació per aconseguir el màxim rendiment. La plataforma haurà de maximitzar la delegació de consultes a les fonts de dades per aconseguir un bon rendiment sense haver de moure les dades a la plataforma de virtualització o copiar les dades a un repositori central.
2. Heu de proporcionar un optimitzador dinàmic basat en costos que seleccioni el millor pla d'execució per a cada consulta, sense dependre d'ajustaments manuals per part dels desenvolupadors.
3. Heu de disposar de tècniques intel·ligents d'acceleració de consultes, basades en IA, per a escenaris analítics complexos. Ha de suportar aggregation awareness, execució paral·lela (MPP).

2.6.2. Detall de Requisits

1. Ha de proporcionar un optimitzador basat en costos (CBO) multifont
2. Heu de disposar de tècniques d'optimització de consultes d'última generació (optimització de predicats, reescriptura de consultes) que incloguin, entre d'altres, l'agregation push-down, join reordering, branch pruning, etc.
3. Heu de disposar d'optimitzacions avançades per bregar amb escenaris de Logical Data Warehouse/Data Fabric. L'optimitzador ha de disposar de tècniques específiques per a consultes sobre grans volums de dades a fonts distribuïdes.
4. Heu de disposar de simplificació automàtica de consultes basades en metadades (per exemple, particions virtuals)
5. Heu de permetre el moviment automàtic de dades com a taules temporals per a l'acceleració de consultes. Utilització de mecanismes nadius de càrrega massiva a la destinació. L'eina ha de suportar la creació de taules temporals a les fonts de dades, la memòria cau o en un motor MPP per accelerar les consultes.
6. Heu de disposar d'integració amb motors MPP per acceleració de consultes en escenaris complexos de Big Data.
7. Heu de disposar d'optimitzacions avançades per aprofitar taules que hi ha a més d'una font de dades (fonts alternatives en l'execució d'una consulta)

8. Heu de suportar conjunts de dades agregades prèviament (també conegudes com a resums o aggregation awareness) per a l'acceleració de consultes. Els resums es gestionen de manera transparent per als usuaris finals
9. Heu de possibilitar les càrregues de forma incremental. La solució ha de suportar operacions UPSERT, és a dir, insereix les noves files i actualitza les files que ja existeixen
10. Heu de permetre la càrrega de resums directament des de la font de dades a la destinació quan tots dos són la mateixa base de dades, és a dir, sense passar per la capa de virtualització de dades.
11. Heu de suportar l'emmagatzematge de conjunts de dades intermèdies en taules de bases de dades externes (taules remotes) per a l'ús posterior. Per exemple, per simplificar els fluxos de científics de dades
12. La plataforma ha de mostrar un pla detallat d'execució per a l'anàlisi del rendiment
13. L'eina ha de permetre assignar recursos d'execució a usuaris/aplicacions segons SLAs. En les regles es poden tenir en compte diferents paràmetres de la sessió (usuari, rol, client, IP d'origen, hora del dia, etc.)
14. Automatització basada en IA: L'eina ha de poder utilitzar les dades d'ús de consultes anteriors per optimitzar el rendiment de consultes posteriors
15. Motor MPP integrat. L'eina ha d'incorporar el motor MPP propi per a un accés eficient als repositoris d'emmagatzematge d'objectes, l'emmagatzematge en memòria cau i l'acceleració de consultes.
16. L'eina ha de proporcionar mecanismes específics per optimitzar/minimitzar el trànsit a través de connexions WAN

2.7. Memòria cau / Replicació Selectiva

2.7.1. Requisits generals

1. L'eina ha de proporcionar capacitats de replicació selectiva de dades per donar resposta als diversos casos d'ús d'un Data Fabric, inclòs l'emmagatzematge en memòria cau, taules temporals i taules permanents. Possibilitat de fer servir aquests mecanismes tant en els orígens de dades com en altres sistemes (per exemple, un data warehouse o un data lake).
2. L'eina ha de proporcionar un suport versàtil per a emmagatzematge en memòria cau, amb un doble propòsit: millora el rendiment ja que les dades es recuperen directament des de la memòria cau en comptes de des dels orígens de dades; evitar accedir als orígens de dades, que podrien estar sobrecarregats en moments de pic de treball.

2.7.2. Detall de Requisits

1. Heu de permetre la replicació de dades per suportar casos d'ús de Data Fabric. Possibilitat de creació de taules temporals o remotes a les fonts, cache, o base de dades externa.
2. Heu de permetre modificar les sentències DDL per crear les taules de memòria cau, taules temporals i taules remotes, per incloure settings concrets de la base de dades de destinació.
3. Ha de tenir la capacitat d'usar vistes com a fluxos ELT per a la replicació directa dins d'una font (per exemple, llac de dades).
4. Heu de donar suport a actualitzacions incrementals (cache, taules remotes, resums, etc.). La solució suporta operacions UPSERT, és a dir, insereix les noves files i actualitza les files que ja existeixen.
5. Heu de permetre habilitar la memòria cau per a una visualització concreta
6. Ha de donar suport per a diferents tipus de bases de dades com motors d'emmagatzematge en memòria cau (en memòria, MPP, núvol, etc.)
7. Heu d'utilitzar els mecanismes de càrrega massiva nadius de la base de dades per a un emmagatzematge en memòria cau més ràpid
8. Ha de tenir la capacitat de carregar dades directament a la memòria cau des de la font de dades quan tots dos són iguals, és a dir, sense passar per la capa de virtualització de dades
9. Ha de disposar de suport per a consultes híbrides, és a dir que tornin les dades contingudes a la memòria cau més les dades que falten ("delta") obtingudes en temps real dels orígens de dades
10. Heu de disposar de la capacitat d'emmagatzemar en memòria cau fonts de dades incompletes (per exemple, serveis web amb paràmetres d'entrada). Suport per a emmagatzematge en memòria cau parcial (és a dir, emmagatzemar en memòria cau només un subconjunt de les dades a l'origen)
11. Heu de suportar el moviment massiu de dades a un emmagatzematge físic per satisfer les necessitats de migració de dades
12. Heu de suportar l'actualització de dades a la memòria cau basades en esdeveniments (JMS o Kafka)

2.8. BigData / Cloud

2.8.1. Requisits generals

1. La plataforma ha de donar suport complet i integrat per a fonts natives al núvol, tant com a fonts de dades, memòria cau o destinació de la replicació selectiva.

2.8.2. Detall de Requisits

1. Ha de proporcionar connectors per a llac de dades, incloses, entre d'altres, Hive, Obligatori: Impala, Presto, SparkSQL i Athena.
2. Heu de proporcionar connectors per a fitxers emmagatzemats al sistema de fitxers Hadoop o en emmagatzematge d'objectes en núvol, inclosos, entre d'altres, Obligatori: HDFS, S3, Azure Data Lake Storage, Azure Blob Storage, Azure Data Lake Storage Gen 2 i Google Cloud Storage
3. Ha de proporcionar connectors per a bases de dades NoSQL orientades a columnes a Hadoop, inclosos, entre d'altres, HBase o MapR
4. Ha de donar suport com a memòria cau de motors Hadoop SQL i magatzems de dades al núvol, inclosos, entre d'altres, Hive, Obligatori: Impala, Presto, SparkSQL, Athena, Redshift, Databricks, Snowflake i Yellowbrick
5. Ha de permetre la possibilitat d'utilitzar motors MPP ja existents al client, local o núvol, per acceleració de consultes

2.9. Semàntica i Governança de Dades

2.8.1. Requisits generals

1. La plataforma ha de disposar de capacitats de modelatge semàntic, inclosos assistents gràfics fàcils de fer servir, captura de llinatge de dades, detecció de canvis i anàlisi d'impacte.

2.8.2. Detall de Requisits

1. Semàntica: El producte ha d'admetre afegir informació semàntica (etiquetes, descripcions) a les vistes i columnes dins d'una vista.
2. Semàntica: El producte ha de suportar la definició de regles de seguretat, inclòs l'emascament, en funció de la semàntica (etiquetes)

3. Governança: L'eina ha de permetre als usuaris recolzar, advertir o alertar altres usuaris sobre certs conjunts de dades
4. Governança: La plataforma ha de mostrar el llinatge de dades des de cada columna d'una vista o entitat de negoci fins als orígens de dades, mostrant totes les transformacions que es fan pel camí
5. Governança: La plataforma ha de detectar canvis a les fonts. La solució ha de suportar l'actualització de metadades d'origen, l'anàlisi d'impacte i la propagació automàtica de canvis als orígens de dades cap als models creats a la plataforma de virtualització de dades
6. Governança: L'eina ha de proporcionar anàlisi d'impacte de modificacions fetes en vistes intermèdies dins del model sobre altres vistes.
7. Integració: L'eina ha de permetre exposar les metadades via API per integrar-les a altres eines de Governança com Colibra, IGC, IMM, Talend
8. Heu de permetre la integració amb eines de Data Quality
9. Heu de permetre la integració amb eines de Data Profiling

2.10. Processament per lots

2.10.1. Requisits generals

1. L'eina haurà de proporcionar capacitats per a l'execució planificada de tasques habituals com a càrrega en memòria cau, replicació selectiva, monitorització de canvis a les fonts, etc.

2.10.2. Detall de Requisits

1. L'eina haurà de suportar la programació de tasques batch per a diferents propòsits.
2. Heu de permetre tasques de càrrega/actualització de memòria cau
3. Heu de permetre tasques d'indexació de contingut, per cerca basada en paraules clau
4. Heu de permetre tasques de gestió de metadades: alertes de canvi de font, recopilació d'estadístiques, manteniment de memòria cau (esborrament automàtic de dades obsoletes), etc.

5. Heu de permetre l'exportació de dades a fitxers o bases de dades externes
6. Integració: la solució ha d'oferir un API per poder invocar l'execució de tasques des d'eines de planificació de tercers, per exemple Control-M
7. L'eina ha de permetre enviar notificacions/alertes per correu electrònic sobre l'execució de tasques
8. L'eina ha de suportar la definició de dependències entre tasques i re intents de tasques

2.11. Desenvolupament i Operacions (DevOps)

2.11.1. Requisits generals

1. La plataforma ha de proporcionar una consola centralitzada per a la gestió de desplegaments complets de la solució, inclosos els entorns de desenvolupament, preproducció i producció.
2. Heu de proporcionar una monitorització en temps real de tots els components de forma integrada.
3. Cal auditar totes les consultes, parametritzacions i activitats realitzades al sistema.
4. S'ha de poder integrar amb solucions de monitorització de tercers.

2.11.2. Detall de Requisits

1. La interfície ha de ser 100% web
2. Desenvolupament: La interfície ha d'incloure tots els passos de la creació d'un model de dades lògiques (des de la connexió a les fonts de dades, la combinació i la transformació d'aquestes per crear vistes lògiques i finalment la publicació de les mateixes per accedir-hi). en múltiples formats)
3. Desenvolupament: L'eina ha de fer una documentació automàtica de serveis de dades. Heu de proporcionar documentació OpenAPI (Swagger) dels serveis de dades publicats. Heu d'incloure un visor web on veure el document OpenAPI i provar les operacions des del mateix visor.
4. Operació: El producte ha de proporcionar una consola central des d'on administrar els diferents entorns (desenvolupament, preproducció i producció) i eines.
5. Operació: El producte ha de proporcionar una gestió centralitzada de llicències

6. Operació: Heu de permetre la distribució centralitzada d'actualitzacions de programari
7. Operació: Heu de permetre la gestió centralitzada de logs generats per tots els servidors d'un clúster
8. Autenticació: Autenticació centralitzada a totes les eines DevOps. Suport SSO
9. Autorització: Diferents rols per a diferents perfils d'usuari, inclosos, entre d'altres, administrador de desplegaments (creació de clústers, instal·lació d'actualitzacions, administració de llicències), administrador de promocions (creació i desplegament de revisions als diferents entorns), administradors de sistemes (modificació dels nivells de log)
10. DevOps: Heu de disposar de suport per a promocions sense interrupció de servei. L'eina s'ha d'ocupar automàticament de gestionar activitats complexes com el balanceig de càrrega entre els nodes mentre es fa la promoció.
11. Automatització: Gestió automatitzada d'infraestructura al núvol. Automatització de la instal·lació, configuració, implementació i actualització de la solució en entorns cloud (AWS, Azure i GCP)
12. Integració: Heu de disposar d'APIs per a la integració amb sistemes de gestió del cicle de vida de tercers, per exemple, Jenkins.
13. Monitorització: Ha de proporcionar una monitorització web integrada, incloent paràmetres de CPU, memòria, sessions, consultes, memòria cau, etc.
14. Monitorització: Els logs s'han de poder desar en disc o en base de dades, a les instal·lacions de TMB o al núvol (Amazon S3 i Azure Blob Storage), per a la seva posterior anàlisi i integració amb eines d'anàlisi de logs de tercers, per exemple, Splunk
15. Monitorització: Possibilitat de carregar logs per "retrocedir en el temps" i representar gràficament l'estat del sistema en un moment donat
16. Monitorització: La informació de monitorització s'ha d'exposar a través de JMX per a la integració amb eines de monitorització de tercers. Integració amb serveis de monitorització al núvol, inclosos, entre d'altres, AWS CloudWatch i Azure Monitor
17. Auditoria: Hi ha d'haver un registre complet de totes les consultes i activitats realitzades al sistema, incloent-hi usuari, IP d'origen, hora, consulta, taules involucrades, etc.
18. Auditoria: Cal registrar els intents d'inici de sessió (exitosos i fallits)
19. Desenvolupament: S'ha d'integrar nadiu amb sistemes de control de versions (SVN, TFS i Git) per a desenvolupament col·laboratiu

2.12. Desplegaments

2.12.1. Requisits generals

1. Ha de suportar desplegaments a les instal·lacions de TMB (on-premise), al núvol públic/privat (als principals hyperscalers, AWS, Azure o GCP), així com a desplegaments híbrids (on-premise + núvol).
2. Ha de suportar desplegaments en contenidors
3. El programari haurà de proporcionar les mateixes funcionalitats al núvol públic/privat i a instal·lacions on-premise.
4. La llicència ha de ser independent del desplegament on-premise, núvol o híbrid, sense conversió de llicències ni costos afegits per canvi de plataforma.
5. Les llicències s'han de basar en el nombre de cors del motor de virtualització. L'ús no ha d'estar restringit pel nombre de consumidors connectats, pel nombre de connexions a orígens de dades (ie nombre de fonts), ni pel volum de dades processades pel motor de virtualització.
6. La solució ha d'incloure un motor MPP totalment integrat a la plataforma.

2.12.2. Detall de Requisits

1. Ha de suportar desplegaments en contenidors (p. ex., Docker), inclosa la compatibilitat amb Kubernetes (K8s)
2. Heu de suportar desplegaments al núvol, incloent-hi AWS, Azure i GCP.
3. Ha de suportar desplegaments híbrids i multinivell
4. Ha de tenir compatibilitat amb desplegaments en contenidors al núvol, inclòs EKS (Amazon Elastic Kubernetes Service), AKS (Azure Kubernetes Service) i GKE (Google Kubernetes Engine)
5. Ha de suportar escalat automàtic
6. Heu de suportar el desplegament directe des dels marketplaces d'AWS, Azure i Google Cloud

3. REQUISITS NO FUNCIONALS

A continuació, es descriuen amb més detall els requisits no funcionals que ha de proporcionar la solució:

1. Ha de ser una solució escalable (horitzontal, vertical)
2. Ha de suportar alta disponibilitat
3. Ha de disposar de còpies de seguretat i recuperació davant de desastres
4. Heu de permetre el xifratge de metadades
5. Heu de disposar d'UI per a diferents rols (administradors, desenvolupadors, administradors de dades, usuaris comercials, etc.)
6. Ha de ser una solució integrada. Amb un esforç mínim d'integració
7. Ha de tenir suport multilingüe
8. Les integracions han de fer servir estàndards oberts
9. Les integracions han d'exposar-ne la funcionalitat en forma d'API
10. Hi ha d'haver una visió i roadmap de producte: Solució de Virtualització de Dades completament dissenyada, desenvolupada i evolucionada per tal propòsit independentment del client/cas específic d'ús, és a dir, haurà de ser una solució COTS (Commercial off the shelf) i no una solució feta a mida per a un client/cas d'ús o desenvolupada a partir de components de tercers.
11. Motor MPP integrat: El proveïdor haurà de proporcionar suport d'aquest motor inclòs a la llicència de suport del motor de virtualització.
12. Motor MPP integrat: Es podrà desplegar fins a 128 vegades el nombre de cors llicenciats pel motor de virtualització.

4. SERVEIS A CONTRACTAR

4.1. Subscripció

Se sol·licita la contractació de la següent subscripció:

- **1 subscripció de Denodo ENTERPRISE SINGLE SERVER 4 Cores 3 years**

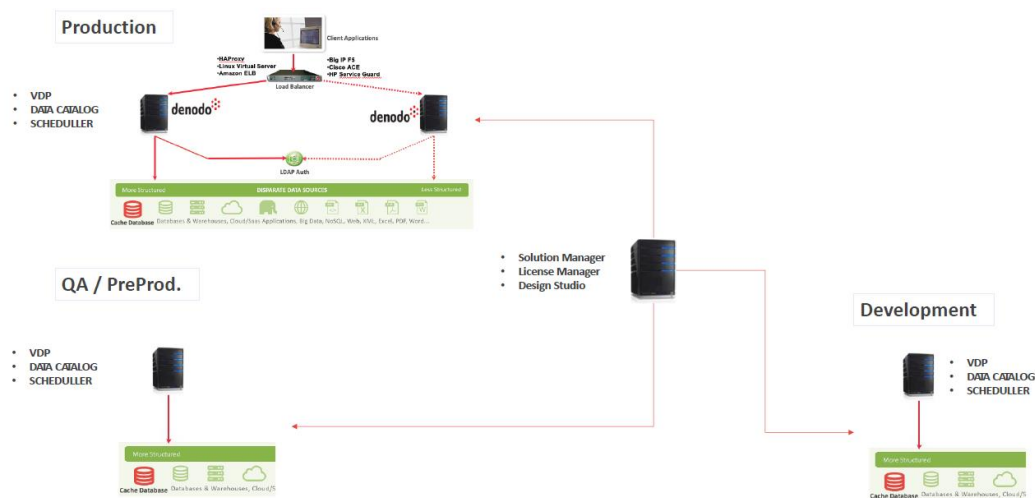
La subscripció ja inclou el suport tècnic següent:

- Suport Premium:
 - Cobertura de suport 24x7 (per a sol·licituds de servei de gravetat 1)
 - Fins a 8 contactes amb clients
 - <= 2 hores de temps de resposta (gravetat 1)
- Formació il·limitada sota demanda
- Accés a totes les classes en línia: arquitectes, desenvolupadors, administradors, usuaris comercials.
- Sessions d'assessorament (DPASS Base Services)
 - Les sessions d'assessorament virtuals de Denodo ajudaran els clients a minimitzar sorpreses i riscos en abordar contínuament els problemes de la plataforma i la identificació d'àrees de millora.
- Avaluacions de la plataforma (DPASS Base Services)
 - Les avaluacions de la plataforma de Denodo ajudaran a identificar i documentar els riscos actuals als sistemes instal·lats, assegurant optimització de recursos i processos, i recerca d'àrees per a millora i creixement general de la solució de dades.

4.2. Servei d'Instal·lació

Se sol·licita el servei d'instal·lació de la plataforma a 3 entorns (Producció, Desenvolupament i Solution Manager):

Denodo Full Architecture



Producció: El sistema de producció s'instal·larà sobre la plataforma Vmware per garantir-ne l'alta disponibilitat (CPD Sagrera).

El HW serà un Servidor Virtual amb 4 Cores dedicat d'ús exclusiu i 32GB Memòria RAM d'ús exclusiu.

Desenvolupament: El sistema de desenvolupament s'instal·larà sobre la plataforma Vmware (CPD Triangle).

El HW serà un Servidor Virtual amb 4 Cores dedicat d'ús exclusiu i 32GB Memòria RAM d'ús exclusiu.

Solution Manager: S'instal·larà de forma independent el Solution Manager que a més acollirà la instal·lació de productiu (VDP, DATA CATALOG i SCHEDULER) per tenir un backup davant de desastres (CPD Triangle).

El HW serà un Servidor Virtual amb 4 Cores dedicat d'ús exclusiu i 32GB Memòria RAM d'ús exclusiu.

4.3. Servei de desenvolupament

Se sol·licita l'adquisició d'una borsa d'hores de tècnics experts a Denodo amb l'objectiu de fer les tasques de configuració i desenvolupament dels diferents projectes que s'aniran construint sobre la plataforma.

Aquesta borsa d'hores s'utilitzaran en format a demanda i només es facturaran les hores consumides, i no serà obligatori per part de TMB el consum total.

El volum d'hores que cal contractar és de 300 hores per any.

5. SERVEI

5.1. RESPONSABILITAT DE L'OFERTANT

L'ofertant es compromet a disposar de personal de servei qualificat i estable per complir les seves obligacions acordades en aquesta contractació. Hauran d'assegurar en tot moment els coneixements sol·licitats i de client i es farà càrrec dels traspassos de coneixement necessaris i específics de client durant la durada del contracte.

5.2. LLOC DE REALITZACIÓ DELS TREBALLS

Les tasques corresponents al projecte es podran realitzar a les dependències de TMB o a les del proveïdor. En cas d'optar per les dependències del proveïdor haureu d'assegurar una comunicació fluida assignant de manera permanent un telèfon i un sistema de missatgeria instantània, així com vídeo conferència quan pugui ser requerida.

5.3. SEGUIMENT DEL SERVEI

Prèviament a l'inici de la contractació es definiran els protocols de seguiment i reporting que TMB vulgui aplicar per a aquest servei.

Es designarà un coordinador màxim del servei per part de l'ofertant i de TMB per resoldre qualsevol tema que afecti el funcionament correcte del servei contractat. Es preveu fer reunions setmanals de seguiment amb la finalitat de garantir i millorar els nivells de qualitat.

Es farà arribar a TMB mensualment una relació de totes les hores consumides per cadascun dels perfils contractats.

Així mateix, es farà un seguiment continu del servei per verificar-ne el funcionament.

5.4. IDIOMA

L'adjudicatari haurà de desenvolupar els sistemes, les interfícies d'usuari i la documentació de les aplicacions en català.

5.5. GARANTIA

Cal garantir l'equip i el coneixement establerts en aquest document.

Si en qualsevol moment en el transcurs de l'execució del servei, el seguiment continuat indica que el servei no és satisfactori, es rescindirà el contracte.

5.6. SEGURETAT I CONFIDENCIALITAT

5.6.1. Confidencialitat i publicitat del servei

L'adjudicatari està obligat a guardar secret respecte de les dades o la informació prèvia que no sent públics o notoris estiguin relacionats amb l'objecte del contracte.

Qualsevol comunicat de premsa o d'inserció als mitjans de comunicació que el proveïdor faci referent al servei que presta a TMB ho haurà d'aprovar prèviament el client.

5.6.2. Propietat intel·lectual

TMB adquirirà en exclusiva la propietat intel·lectual de tot el material que sigui elaborat per l'adjudicatari en execució del contracte i en particular, de tots els drets de propietat intel·lectual patrimonial, industrial i d'imatge que se'n derivin inclosa l'explotació en qualsevol modalitat i baix qualsevol format, per a tothom, del treball elaborat per l'adjudicatari o els seus empleats en execució del contracte, reservant-se TMB qualsevol altra facultat annexa al dret de propietat intel·lectual o industrial.

Serà propietat de TMB el resultat dels serveis així quants materials i documents es realitzin en compliment del contracte.

TMB serà titular de tots els drets referits al paràgraf anterior pel termini màxim legal permès i l'única organització que per aquest concepte podrà explotar i comerciar amb el treball desenvolupat en execució del Contracte, abans o després de la seva finalització, corresponent als autors materials únicament els drets morals que els reconeix la legislació vigent en matèria de propietat intel·lectual.

Als efectes previstos als dos paràgrafs anteriors, l'adjudicatari es compromet al lliurament de tota la documentació funcional i tècnica, així com materials i entregables generats durant la prestació del servei i en el procés d'anàlisi, disseny, desenvolupament, implantació i proves de les mateixes. Tota la documentació elaborada i els resultats obtinguts per l'adjudicatari en execució del contracte seran propietat de TMB, al poder quedaran a la finalització del contracte, no podent l'adjudicatari utilitzar-la per a altres persones, entitats o empreses.

L'adjudicatari respondrà de l'exercici pacífic de TMB en la utilització del programari i altres drets proporcionats per l'adjudicatari amb motiu del contracte i serà responsable de tota reclamació que pugui presentar un tercer per aquests conceptes contra TMB i haurà d'indemnitzar TMB per tots els danys i perjudicis que aquesta pugui patir per aquesta causa. En tot cas, les relacions jurídiques derivades del contracte s'establiran entre TMB i l'adjudicatari. TMB no estarà contractualment vinculada a persones diferents de l'adjudicatari.

5.6.3. Seguretat i protecció de dades

L'adjudicatari dels serveis es compromet a complir els requisits de seguretat, de protecció de dades i de continuïtat aplicables a fi del contracte especificat a:

- La legislació vigent en general i, en particular, quan es tractin dades de caràcter personal, la Llei Orgànica 3/2018 de Protecció de Dades Personals i garantia dels drets digitals i El Reglament General de Protecció de Dades (RGPD).

Adicionalment, l'adjudicatari es compromet a:

- Complir les directives tecnològiques, de seguretat i de qualitat que estableixi el client.
- Implementar les mesures, els processos, i els requeriments que el client sol·liciti a tal fi i li proposarà els que consideri necessaris per millorar les solucions.
- Facilitat tota aquella informació que el client requereixi perquè pugui donar compliment a la legislació i la normativa referida en aquest apartat.

Per motius de garantir la seguretat, qualsevol compartició de recursos tècnics (infraestructura, maquinari, etc.) utilitzats en el marc de l'execució del contracte serà prèviament justificada al client amb un informe d'anàlisi de beneficis i de riscos que aquest haurà d'aprovar.

Els adjudicataris utilitzaran la xarxa, el maquinari i/o programari propietat de TMB exclusivament per a l'ús o el benefici de TMB.

Carles Teixit Galvez
Responsable Projectes Grup Prestació del Servei