

Enpresa-Garapenerako Euskal Agentzia  
Agencia Vasca de Desarrollo Empresarial  
Business Development Basque Agency

[www.spri.eus](http://www.spri.eus)



SPRI-ENPRESA GARAPENERAKO EUSKAL AGENTZIA

PRESKRIPZIO TEKNIKOEN AGIRIA

**DATUAK BIRTUALIZATZEKO PLATAFORMA**

## AURKIBIDEA

|                                       |           |
|---------------------------------------|-----------|
| <b>1. KONTRATUAREN XEDEA</b>          | <b>3</b>  |
| <b>2. ZEHAZTAPEN TEKNIKOAK</b>        | <b>5</b>  |
| 2.1. KONEKTIBITATEA                   | 5         |
| 2.1.1. Konexio motak eta integrazioak | 5         |
| 2.1.2. Datu-iturriak                  | 6         |
| 2.2. DATUAK                           | 7         |
| 2.2.1. Modelatzea                     | 7         |
| 2.2.2. Entregatzea                    | 9         |
| 2.2.3. Katalogoa                      | 10        |
| 2.2.4. Semantika                      | 10        |
| 2.2.5. Cachea                         | 11        |
| 2.3. SEGURTASUNA ETA GOBERNANTZA      | 12        |
| 2.4. OPTIMIZAZIOA ETA ERRENDIMENDUA   | 14        |
| 2.5. CLOUD                            | 17        |
| 2.6. LOTEKA PROZESATZEA               | 17        |
| 2.7. DEVOPS FUNTZIOAK                 | 18        |
| 2.8. HEDATZEAK                        | 18        |
| <b>3. BESTE ESPEZIFIKAZIO BATZUK</b>  | <b>19</b> |

## 1. KONTRATUAREN XEDEA

SPRI-Enpresa Garapenerako Euskal Agentzia (aurrerantzean, “SPRI”) Ekonomiaren Garapen, Jasangarritasun eta Ingurumen Sailaren mendeko zuzenbide pribatuko erakunde publikoa da, eta Euskal Autonomia Erkidegoko enpresei laguntze aldera tresna mota desberdinak kudeatzen dihardu, enpresen lehiakortasuna hobetzen eta eskualdea ekonomiaren ikuspegitik garatzen laguntzeko. Oro har, tresna horiek laguntzen programa edo zerbitzuen modura gauzatzen dira enpresa-esparru desberdinetarako, hala nola ekintzailatza, digitalizazioa, I+Ga, berrikuntza edo industria-garapena.

Jarduera garatzeko, SPRIk zenbait aplikazio eta zerbitzu erabiltzen ditu, eta datu-bolumen desberdinetan oinarritzen dira. Datuen gordailuak gaur egun deszentralizatuta daude, beraz, eskuratzea zailtzen dute, eta informaziora sartzeko puntu logiko bakarra izatea ezinezkoa da.

Kontratuaren xedea da datuak birtualizatzeko plataforma ezartzea. Horri esker, SPRIk datu-iturrien heterogeneotasunaren hesia deuseztatuko du. Birtual bihurtzearen bidez, barneko eta kanpoko datuetarako sarbidea homogeneoa, bateratua eta gidatua izango da, eta oraingo nahiz etorkizuneko eragiketa- eta informazio-beharrak zein behar teknikoak aseko dira.

SPRIk, datuak birtualizatuta ase nahi dituen beharren artean, nahi duten aplikazio/zerbitzu guztiei puntu logiko bakarra emango die, irisgarritasuna eta eskuragarritasuna sinplifikatzeko, puntuz puntuko konexio konplexuak eta mantentzeko zailak saihesteaz gain. Helburu hori betetzen denean, nahi izanez gero, kontsumitzaileei datuak gordetzen diren iturrietatik banantzea lortuko da.

Datuak birtualizatzeko geruza hori dela bide, datu fisikoei buruzko negozio-ikuspegi ugari sortuko lirateke erabiltzaile eta/edo aplikazio/zerbitzu bakoitzaren berariazko eskakizunetara egokitzeko, hau da, izendapen-konbentzioak eta berariazko formatu nahiz eskema. Datuak eskura jartzeko orduan, beharrezkoa izango da birtualizazio-geruzak emate-konponbideak eskaintzea hala eskatzen duten aplikazioen eta zerbitzuen beharrak asetzeko (adibidez, sarbideak: JDBC/ODBC, REST/JSON, etab.). Hala, SPRIk datu-ikuspegi desberdinak izango ditu, gordailuak bikoiztu gabe eta gordailu zentralizatu batean kopiatu behar izan gabe. Modu horretan, bere baliabideak optimizatuko ditu.

Datuak birtualizatuta, SPRIk bere aplikazioen, zerbitzuen eta erabiltzaileen bidez datuetara sartzeko funtzionaltasun berriak izango ditu. Baina, gainera, banantze-geruza bat sortuko du datu-baseen, aplikazioen eta zerbitzuen bilakaera errazteko. Xede horrez, datuen geruzaren eta gainerako sistemen artean pantaila bat ezarriko da (hala, adibidez, azpiegitura batetik datu-basea migratu ahaliko da, konexio-kate berdina mantenduta eta datu-base horrekin lotutako zerbitzuak gelditu gabe).

Aurrekoan funtsatuta, SPRIk birtualizazio-plataforma izango du identifikatu diren honako helburu hauei erantzuteko:

- A. **Abstrakzioa.** Behar duten zerbitzu eta aplikazio guztiek sarbide-puntu logiko bakarraren bidez datuetan sartzeko aukera izango dute. Puntu horrek teknologien eta ereduaren konplexutasuna kontsumitzaileetatik bereiziko du.
- B. **Geruza semantiko bateratua.** Tresnan izendatutako erabiltzaileek negozio-entitateak definitu ahalko dizute, eta SPRI guztirako definizio erkidea izango da. Horrez gain, entitate horiek datu-zerbitzu gisa erakutsi ahal izango dira aplikazioek eta zerbitzuen erabiltzaileek kontsumitzeko.
- C. **Datuen federazioa.** Plataformak zenbait jatorriko datuak konbinatzen utziko du datuak birtualizatzeko geruzaren bidez sartzeko, gordailu zentralizatu batean kopiatzeko beharrik izan gabe.
- D. **Prozesatzea.** Plataformak gehienez aprobatutako datuen jatorrien prozesatze-gaitasunak (datu-bolumen handien gainean eraginkortasunez baturak egitea, hodeian ostatatuta dauden datuen gainean kontsultak gauzatzea, etab.). Gainera, birtualizazio-plataformak birtual bihurtu diren datuen geruza osatzeko orduan, sarearen bidez datuen joan-etorriak minimizatuko ditu.
- E. **Aldaketak kudeatzea.** Datu-iturrietan edo azpiko teknologietan aldaketak geruza horretan erraz egokitu ahalko dira, eta aplikazio kontsumitzaileen gainean eragina minimizatuko da. Hala, SPRIk gutxienez bere sistemak osagai teknologiko berriekin bilakatzeko aukera izango du.
- F. **Segurtasun-geruza eta gobernantza bateratua.** Plataformak puntu bakarretik datuen segurtasun- eta gobernantza-arauak definitu, administratu eta betearazi eta, gainera, erakundeko eredu kanonikoari aplikatu ahal izango ditu (negozio-entitateen motak, elementuak, atributuak eta erlazioak).
- G. **Erabiltzeko erraztasuna.** Konponbidea erraz erabiliko da. Datuen iturrietara konektatzeko, negozio-entitateak sortze aldera konbinatzeko eta protokolo anitzen bidez argitaratzeko prozesua modu grafikoan egin ahalko da. Gainera, negozio-erabiltzaileei laguntzeko metadatuaren katalogoa izango da sarbidea errazago kudeatzeko.
- H. **Malgutasuna.** Plataforma malgua izango da eredu berriak sartzeko eta negozio-beharrei arin egokitzeko.
- I. **Errendimendua.** Plataformak errendimendu ezin hobea emango du eta, horretarako, optimizazio-teknika aurreratuak konbinatuko dira, eta cacheak erabiltzeko aukera izango du. Horrez gain, kostuetan oinarritutako optimizazio-motorra izango du, egikaritze-denboran, ahal diren artean automatikoki estrategiarik onena aukeratzeko, ad-hoc konfigurazioaren mende egon gabe.

- J. Negozio-erabiltzailearentzako ezaugarriak.** Plataformak erakundeko datuen kontsumitzaile bakoitzari arin eta erraz emango dio behar duen informazioa berak behar dituen formatuetan eta, aldi berean, datuen koherentzia, gobernantza, kalitatea eta segurtasuna bermatuta.

## 2. ZEHAZTAPEN TEKNIKOAK

### 2.1. KONEKTIBILITATEA

Gaur egun, SPRIk plataformen sistema heterogenoa du, eta egunero erabiltzen dituen datuak (datu-baseak, fitxategien sistemak, baliabide partekatuak, aplikazioak, zerbitzuak, etab.) horietan ageri dira. Enpresaren barruan datuen aktiboak sakabanatuta izatearen ondorioz, batzuetan datuak erreplikatu, kopiatu, etab. egin behar dira. Arazo honi aurre egite aldera, birtualizazio-plataforma honako hauek gauzatzeko gai izango da:

- A. Datu-iturri guztien abstrakzio-geruza osoa ematea.** Sarbide-puntu bakarra sortuko da edozein datu-multzora sartzeko eta datuak ez kopiatzeko/errepikatzeke eta, bide batez, horien ondoriozko mantentze-lanak saihesteko.
- B. SPRIren datu-aktibo guztietarako konexioa ahalbidetzea.** Fabrikako konektoreen zerrenda konektore pertsonalizatuekin (ad-hoc) zabaldu ahal izateaz gain.
- C. Prozesatze-gaitasunak maximizatzen dituzten konektoreak erabiltzea.**

#### 2.1.1. Konexio motak eta integrazioak

Ondoren, datuak birtualizatzeko plataformaren bidez bete beharko diren integrazioak zehazten dira:

- A. Jatorrizko integrazioa** beste fabrikatzaile edo zerbitzu batek emandako kontroladoreak eta APIak erabilita.
- B. Jatorrizko integrazioa datu-iturriekin hodeian** (Snowflake, Redshift, BigQuery, Presto, Synapse, etab.). Gainera, konexioak egin ahalko dira, bai zuzenean, bai cachea sortuta.
- C. Konektore pertsonalizatuak garatzeko SDK.**
- D. Datu-iturrien jatorrizko APIak erabiltzea** zutabeen eta bestelako metadatuak deskriptoreak lortu eta integratzeko, bai eta errendimendua hobetze aldera erabil daitezkeen estatistikak edo indizeak lortzeko ere.

Gainera, plataformak datu-iturrien kargako jatorrizko APIak erabili beharko ditu (BULK data load APIak) (bai irakurketan, bai idaztean cachearen kasuan) goi-errendimendua emateko.

- E. Datu-iturrien mota desberdinekin integratzea**, honako hauek barne: datu-base erlazionalak, datu-base paraleloak eta banatuak, dimentsio askoko datu-baseak (SAP BW, etab.), hodeian dauden datu-gordailuak (Amazon, Azure, Redshift, Snowflake, etab.), paketatutako aplikazioak (SAP, NAVISION, SAGE, etab.), SaaS (Salesforce, etab.), web zerbitzuak (SOAP, REST, OData), Excel kalkulu-orriak, XML, JSON, fitxategi lauak, log fitxategiak, JMS buztanak, iturriak streamingen (Kafka, etab.), NoSQL datu-baseak (Cassandra, MongoDB, etab.) eta Hadoop ekosistema oso-osorik (HDFS, Parquet, AVRO, HBase, HIVE, SPARK).

### 2.1.2. Datu-iturriak

Birtualizazio-plataforma zer datu-iturritara konektatu nahi den definitzeko orduan, SPRIk euskarri eman nahi die egun dagoeneko erakundearen azpiegituraren barruan dauden iturriei eta informazio-sistemak bilakatzearen ondorioz sor daitezkeenei. Beraz, besteak beste, honako datu-iturri mota hauei emango zaie euskarri:

- A. Datu-base erlazionalak:** Microsoft SQL Server, ORACLE, DB2, MySQL eta PostgreSQL, etab.
- B. NoSQL datu-baseak:** MongoDB, Cassandra eta Neo4j.
- C. Memoriako datu-baseak:** Redis, Oracle In-Memory, etab.
- D. On Premise DataLake-ak:** Hive, SparkSQL, Impala eta Presto.
- E. Cloud DataWarehouse/DataLake-ak:** Redshift, Snowflake, AWS RDS, Databricks Delta, Azure Synapsis, Google BigQuery.
- F. Fitxategi mugatuak eta egituratuak:** Logen fitxategiak, Excel, XML eta JSON.
- G. Objektu banatuetan dauden gordailuetako fitxategiak:** HDFS, AWS S3, Azure ADLS eta Google Cloud Storage.
- H. SaaS APIak:** Google Analytics, ServiceNow, Microsoft Dynamics 365 eta Twitter.
- I. Web Services APIak** (SOAP, REST, OData)
- J. Streaming/IOT iturriak:** Kafka, JMS, etab.

## K. LDAP eta Active Directory.

## 2.2. DATUAK

### 2.2.1. Modelatzea

Datuak birtualizatzeko plataformak ikuspegi integratuak sortuko ditu edozein iturriko datuak konbinatzeko. Gainera, SPRIren beharrak direla eta, iturri horiek SPRIren azpiegituran daudenak edo bigarren kokapenetan, hala nola hodeietan, hirugarrenen zerbitzuetan, eta abarreetan, daudenak izango dira.

Datuak modelatzeko ingurune lagunkoia emango da, zeinetan tresnaren erabiltzaileek datuen ereduak diseinatu ahalko dituzten, Microsoftek Microsoft SQL Server Management Studioarekin egiten duen bezala.

SPRIk plataforman datuak modelatzeari dagokionez honako eskakizun hauek ase nahi ditu:

- A. Plataforma erabiltzailearen interfaze komuna emango du** diseinatu eta garatzeko. Diseinu-fase guztietarako testuinguruan oinarritutako erabiltzailearen interfaze grafikoak espero dira, datuen konektibitatea, eraldaketa, datuen kalitatea, kontsulten optimizazioa, eraginaren analisia, etab. barne.
- B. Interfazeak zenbait menu eta tresna izango ditu:** Drag and Drop, auto-osatu, laguntza linean, baliozkotzea.
- C. Interfazeak garbitzeko tresnak izango ditu.** Hala, diseinu segurua ziurtatuko da.
- D. Elei esker, erabiltzaileek oso-osorik diseinatu ahalko dute datu-geruza birtuala.** Datu-iturrietara konektatzeko, ikuspegi eratorriak sortze aldera konbinatzeko eta ikuspegiak argitaratzeko prozesua Drag and Drop menuen, laguntzaileen eta ekintzen multzo grafikoaren bidez maneiatuko da.

Erabiltzaileek ikuspegi eratorri mota desberdinak erabili ahalko dituzte. Ikuspegi horiek oinarritzko ikuspegiaren eta bestelako ikuspegi eratorrien arteko konbinazioak/eraldaketak zehaztuko dituzte Join, Intersection, Union, Minus eragiketen eta beste erlazioko aljebra-eragiketen bidez.

Ikuspegi eratorriak grafikoki definitu ahalko dira diseinu-tresnatik edo ANSI SQLrekin bateragarria den birtualizatzeko kontsulta-hizkuntzaren bidez.

- E. Datuen geruzaren diseinuari lotutako baimenak** erabiltzaileak eta rola definituz kudeatu ahal izango dira (bai eskuz, bai SPRIren direktorio aktiboaren bitartez).

- F. % 100 web-garapenaren ingurunea emango da.**

- G. Sintaxia SQLrekin bateragarria.** Datuak modelatzeko interfaze grafikoarekiko komando-sekuentziaren alternatiba emango da. Horretarako, DDL eta DML sententziak erabili ahalko dira.
- H. Datu mota konplexuetarako jatorrizko euskarria.** Euskarri honi esker, egitura hierarkikoak modelatuko dira, hala nola web zerbitzuetako APIak.
- I. Datuen tipologia.** Plataformak datu mota desberdinei euskarria emango die. Oinarrizko motak (Integer, Long, Float, Double, Boolean, DateTime, XML, Bynary, BLOB, etab.) eta mota konplexuak (matrizeak, erregistroak eta enumerazioak, datu geoespazialak, etab.) eutsiko dira.
- J. Metadatuak.** Birtualizazio-plataformak konektatzen den datuen iturrietatik metadatuak eta eskemak automatikoki aurkitzen utziko du. Datuen jatorrietatik ateratako metadatuaren informazioak honako hauek izango ditu: jatorrian datu mota, jatorrizko eta kanpoko gakoak, NOT NULL murrizketak, eremuko indizeak eta estatistikak (zutabe bakoitzeko balio bakarren kopurua, nuluaren kopurua, errenkaden guztizko kopurua, zutabearen batez besteko tamaina, zutabe bakoitzeko gutxieneko eta gehieneko balioak, etab.).
- K. Datuen jatorrizko irudikapena.** Datu konposatu moten jatorrizko irudikapena eutsiko da (adibidez, XMLko datu hierarkikoak, web zerbitzuak, dokumentuetarako datu-baseak, erlazioak sortzen ez dituzten datu-baseak, etab.). Horri esker, egitura hierarkikoak modelatuko dira, adibidez, APIek itzultzen dituztenak.
- L. Testu-balio handietarako eta balio bitarretarako euskarria.** Testu-balio handiak eutsiko ditu, hala nola sarrerako eta irteerako parametroak. Balio bitarrek irudiak, PDF dokumentuak, Microsoft Officeren dokumentuak, testu-dokumentuak, etab. izan ahalko dituzte.
- M. Hirugarrenen modelatze-tresnekin integratzea.** Plataformak industriaren erreferentziazko datuen ereduak/estandarrak eta XSD, WSDL, XML artxiboen eta hirugarrenen modelatze-tresnen eskemen definizioak inportatu ahal izango ditu (besteak beste, CA ERwin, ER/Studio ontziraletakua edo IBM Infosphere Data Architect).
- N. Datuak eraldatzeko funtzioen liburutegiak.** Espero da plataformak halako funtzio hauek betetzea: aritmetikoak, trigonometrikoak, testuak eta datak prozesatzea, motak bihurtzea, funtzio estatistikoak eta enkriptatzekoak, zutabea errenkada eta errenkada zutabe eraldatzea, baldintzapekoak, XML, XPath/XQuery, XSLT, testu-parekotasuna, bilaketak eta mapak egitea, Nest/Flatten-en operadoreak datu hierarkikoak manipulatzeko, adierazpen erregularrak, etab.

- O. Funtzio geospazialen euskarria.** Datuak birtualizatzeko plataformak datu geospazialeterako euskarri osoa emango du, bistaratzeko tresna geospazialean ikuste aldera.
- P. Ikuspegiak eta zutabeei etiketak (“tag”) esleitzea.** Ikuspegiak eta zutabeak deskribatzen laguntze aldera etiketak erabiltzeko aukera emango da. Horri esker, nabigatuz edo bilaketen bidez aurkituko dira, eta katalogoko elementuak SPRIren barruan defini daitezkeen irizpideen arabera (ikuspegi publikoak, pribatuak, etab.) sailkatzeko erabiliko dira.
- Q. Bertsioen kontrol-sistemekin jatorrizko integrazioa.** Plataforma, besteak beste, honako tresna hauekin integratu ahal izango da: Subversion, Team Foundation Server (TFS) edo GIT elkarlaneko garapenerako.

### 2.2.2. Entregatzea

Datuak birtualizatzeko plataformari dagokionez, SPRIk estali nahi dituen zerbitzuen barruan dago Ad-Hoc SPRIrako garatutako aplikazio eta zerbitzuetarako mahaigaineko aplikazioak, zerbitzuak, APIak, etab.) eta merkatuko beste produktu batzuetarako (PowerBI, QlikView, Navision, Salesforce, etab.) birtual bihurtutako datuak entregatzea.

Halako beharrak asetzeko zenbait eskakizun orokor ezartzen dira SPRIren produktuen eta zerbitzuen bilakaera dela eta, egun eta etorkizunean SPRIk berak erantzun behar dituen beharrak betetze aldera.

- A. Datuak entregatzeko interfaze asko.** Plataformak interfazeen bidez datuak sinkronian entregatuko ditu:
- SQL (JDBC, ODBC eta ADO.NET)
  - API (SOAP, REST XML, JSON, HTML edo RSS gisa ordezkatuta)
  - OData
  - GraphQL Web Services
- B. Mezu-ilaren bidez modu asinkronikoan entregatzea.** Hona hemen adibide batzuk: JMS edo Apache KAFKA.
- C. APIak sortu eta hedatzea.** Plataformak APIak sortzen eta argitaratzen utziko du garapen zereginik gabe (Zero-Code). Gainera, SOAP, REST, OData 2 eta OData 4rako euskarria sartuko da.
- APIak sortu eta hedatzeko funtzionalitatearen barruan, SPRIk OpenApi (Swagger) erabiltzen ari den dokumentazioa sortzea bilatuko du.
- D. Aplikazioetarako datuak entregatzea.** Datuak birtualizatzeko plataforma Business Intelligence eta Datuen Analitika aplikazioekin erabili ahal izango da.

Gaur egun SPRIk erabiltzen dituen gain (PowerBI, QlikView, SQL Server Reporting Services, Excel)

- E. GraphQLrako euskarri integratua.** Ikuspegi birtualak GraphQL bidetik hel egiten utziko du. Gainera, plataformak puntu askotatik datuak konbinatu beharko lituzke.

### 2.2.3. Katalogoa

Aurreko puntuarekin batera (alderdi teknikoa aipatzen du), SPRI negozio-erabiltzaileei hurbildu nahi zaie datuak eta autozerbitzua aurkitzeko gaitasunak emate aldera. Hala, Informazio eta Komunikazioen Teknologia Arloarekin elkarrekintzak eta mendekotasunak murriztuko dira eta web tresna bakarrean, modu autonomoan, metadatuak nahiz datuak arakatuko dituzte.

Datuen autozerbitzuaren funtzionalitatean adierazi den helburua lortzeko, plataformak honako alderdi hauek estaliko ditu:

- A. Interfazea 100% web.** Web bat eskuratuko da, eta datuen katalogoa txertatu beharko du SPRIko erabiltzaileek (negozioakoak eta teknikoak) datu eskuragarrien multzoa bilatu eta arakatzeko.
- B. Katalogo sinkronizatua.** Katalogoa datuen plataformarekin sinkronizatuko da, hau da, ikuspegian/kontsultan egiten den edozein aldaketa automatikoki eta berehala datuen katalogoan ere egingo da. Hala, ziurtatuko dugu zerbitzua plataformaren isla leiala izatea.
- C. Estatistikak.** Plataformak datuen katalogoan argitaratutako datu-multzoen erabilera-estatistikak bistaratuko ditu. Halaber, espero da plataforma estatistiketan oinarritutako datu-multzoak iradokitze gai izatea.
- D. Elkarlaneko funtzioak.** Erabiltzaileek datu-multzoak babestu edo haiei buruzko iruzkinak ala ohartarazpenak erregistratu ahalko dituzte. Horri esker, datu-multzoaren erabilera testuinguruan jarriko da, beste erabiltzaile batzuek nola esperimentatzen dituzten ulertuko da, balizko inkongruentziak konponduko dira, etab.

### 2.2.4. Semantika

Semantikaren arloaren barruan, SPRIk gehitzen diren datu guztiak modelatu, horien aldaketak aztertu eta leinua bistaratzea ahalbidetzen duen plataforma izan nahi du. Horretaz gain, plataforma hirugarrenen tresnekin osatu ahal izatea nahi da, plataforman eskaini litekeen geruza semantikoa osatzeko.

Honako hauek dira SPRIk datuen semantikaren arloan birtualizazio-plataformatik lortu nahi dituen gaitasun aipagarrienak:

- A. Modelatze semantikoa.** SPRIri eskaintzen zaion tresnak eredu semantikoa ahalbidetu beharko du. Eredu semantiko hori modu errazean kudeatu ahalko da, laguntzaile grafiko baten bidez.
- B. Gehikuntza semantikoa.** Plataformak ereduei informazio semantikoa gehitzea ahalbidetu beharko du (ikuspegiak, taulak, zutabeak, etab.) etiketak erabiliz. Informazio semantiko hori lagungarria izango da SPRIrentzat, sarbiderako politika eta kontrol orokorrak ezartze aldera (adibidez, etiketa zehatzarekin etiketatutako ikuspegi/zutabe guztiei aplikatuz).
- C. Analisi grafikoa.** Tresnak “datuen leinua” eta “datuen erabilera” arloetako analisi grafiko osoa egiteko aukerak eskainiko ditu. Lortu nahi den leinua hori irudikatuko du, norabide bietan (datuen iturritik negozio-bistetara, eta alderantziz).
- D. Iturrietan aldaketak detektatzea.** SPRIren jarduera-arloen barruan jatorrizko eskemak aldatzen diren kasuak jaso ahal izateko, tresna gai izan beharko da datuen iturriak eguneratzeko eta aldaketen eraginak aztertzeko funtzioak emateko, eredu birtuala berritu ahal izateko.
- E. Tresnekin integratzea.** Metadatuak kudeatzeko soluzioen artean integratzeko aukera. Gaur egun SPRIk ez duenez horrelako tresnarik, metadatueterako hainbat tresnarekin bateragarritasuna maximizatzea gustatuko litzaioke: Collibra DGC (Collibra Data Governance Center), IBM IGC (Information Governance Catalog), IMM (Informatica Metadata Manager), Informatica EDC (Enterprise Data Catalog), Talend Data Catalog, etab.

#### 2.2.5. Cachea

Gaur egun, SPRIk ez du berehala datuetarako sarbidea behar ez duten jarduera horiek egiteko (Real-Time) informaziorako arlorik. Hori dela eta, eta datu-arlo operatiboek eragina izan ez dezaten, informazio hori cachean gordetzeko beharra identifikatu da. Informazioaren cachea erabiliz, lortu nahi dena da denbora errealeko sarbidea behar ez duten zereginetarako operatibo dauden datu-arloen gainean murriztea, besteak beste, honako hauetan: business intelligence txostenak, analitika deskriptibo/prediktiboa, machine learning, etab.

Helburu hori bete ahal izateko, honako hauek dira birtualizazio-plataformarekin estali nahi diren ezaugarriak:

- A. Erreplikazio selektiboa.** Plataformako profil teknikoek datuen erreplikazio selektiboa egitea ahalbidetzeko funtzioaren bidez, SPRIIn ager litekeen edo jada dauden erabilera-kasuei erantzuna ematea lortu nahi da. Gainera, espero da

erreplikazio horrek zenbait cache mota eskaini ahal izatea: mailakatur, partziala, TTLarekin, etab.

**B. Gordetzeko euskarria.** Espero da plataformak euskarri moldakorra izatea gordetzeko, eta, gainera, euskarri hori datuen iturrietatik bereizita egotea. Helburua izango litzateke errendimendua eta datuen iturrien isolaketa hobetzea, datuen iturrietarako sarbidea saihestu ahal izateko, eta, arlo operatiboen kasuan, gaitzargarik ez egoteko. Horretaz gain, euskarriari esker, SPRIk gordetzeko euskarria dimentsionatu ahalko luke, errendimendua hobetze aldera.

**C. Datuak kopiazea (ETL, Data Fabric).** Datuetarako sarbidea izan ahal izateko kanpoko tauletan kopiazu bidez, hala nola eremu operatiboa bereizi nahi denean edo iturritako zuzeneko sarbideak ez duenean errendimendu egokia ematen. SPRIk bilatuko du plataformak datuak “landing” eremu batean kargatu ahal izatea modu errazean (ez du zertan cachearen toki berean egon behar), trantsizioa guztiz arina izan dadin.

Datuak kopiazea erabilgarria izango da kanpoko datu-baseetako tauletan ere bitarteko datuak gordetzeko.

**D. Errendimendua.** Datuen cachea erabiltzearekin errendimendua hobetu nahi da, cache horretako datuak zuzenean gordez. Datuak birtualizatzeko geruzaren cache hori erabiliz, itxaronaldiak laburtuko dira eta errendimendua areagotuko da. Hala, beste moduz datuetarako sarbidea atzeratu edo degradatuko litzatekeen (itxaronaldiak, sistemak gaitzargatzea, etab.) eragiketa konplexuak programatu ahalko dira.

**E. Teknologia anitzak.** Konponbideak ez du cacherako euskarri bakarra eman behar (edo kopuru oso mugatua), teknologia hauetako edozeinetan oinarritutako cache banatu eta mailguan gordetzea jasan behar du:

- Datu-base erlazionalak: Oracle, SQL Server, DB2 edo MySQL.
- Datu-base paraleloak: Teradata, Netezza, Vertica.
- Memoriako datu-baseak: SAP HANA, Oracle TimesTen edo Oracle 12c IM.
- Hodeiko datu-baseak: AWS Redshift edo Azure SQL.
- Prozesatze masibo paraleloko sistemak (MPP): Spark, Hive, Impala edo Presto.

**F. Kokapen fisikoa.** Nahi da cachean gordetzea plataforman bertan ez egitea, konfiguratu daitezkeen kanpoko euskarrietan baizik. Estrategiak baterako kokapena aprobetxatuko du cachean gordetako datuen eta horretatik kanpo baina sisteman bertan gordetzen diren datuen arteko konbinazioak murrizteko (ahal den neurrian).

## 2.3. SEGURTASUNA ETA GOBERNANTZA

- A. Gobernantza- eta segurtasun-geruza bateratua.** Plataformaren bidez, datuen segurtasun eta gobernantza-arauak toki bakarretik zehaztu, kudeatu eta betearazi ahaliko da. Gainera, politika horiek negozio-erakundearen eredu kanoniko baten bidez zehaztuko dira.
- B. Gobernantza eta segurtasunerako euskarri osoa.** Euskarriak datuen maskaratze aurreratua eta roletan oinarritutako sarbide-kontrola (RBAC) izango ditu. Euskarria eskuragarri egongo da erakundeko datuen aktibo guztietarako, kontrolerako eta kudeaketarako puntu bakar baten bidez.
- C. Daturako sarbiderako puntu logiko bakarra.** Datuak birtualizatzeko SPRIri kontsumoko aplikazioetatik informazioaren iturrietara puntutik punturako konexioak saihestea ahalbidetuko dion sarbiderako puntu logiko bakarra eskainiko du.

Birtualizazio-plataformaren erabilera erakundearen datuetarako sarbiderako puntu bakar gisa hedatu ahala, SPRIk plataforma hori zehaztasun-maila oso altua duen eredu kanonikoaren terminoekin zehaztu daitekeen (adib. "Proiekturako" Sarbidea, "Laguntza-programa", "Zenbatekoa", etab.) sarbiderako segurtasun-murrizketak betearazi ahal izateko tresna gisa.

- D. Aplikazio kontsumitzaileetarako sarbidea.** Birtualizazio-plataformak bermatu beharko du aplikazio kontsumitzaileek segurtasuneko konexio eta protokoloak erabiltzen dituztela sarbiderako, hala nola SSL/TLS, HTTPS edo SFTP.
- E. Baimena eta autentifikazioa.** Plataformak eskema-mailako (datu-baseak, taulak, ikuspegiak, etab.) eta datu-mailako (ikuspegi bateko errenkada/zutabeak, taula bateko errenkada/zutabeak, etab.) baimenak dituzten erabiltzaile/roletarako integratutako baimen- eta autentifikatze-mekanismoak onartu beharko ditu.
- F. Eskuordetutako autentifikatzea.** Plataformak autentifikatzea kanpoko LDAP edo Active Directory bat eskuordetu ahal beharko dio, plataformak izan zezakeen kudeaketa-sisteman erabiltzaileak zehaztea beharrezkoa izan ez dadin. SPRIren autentifikatze-sistemak rolak emango dizkie erabiltzaileei, eta, era berean, plataformak horiek erabiliko ditu datuak birtualizatzeko zerbitzariaren barruan edozein datu-base edo bistarako sarbidea murrizteko.
- G. Single sign-on euskarria (SSO).** Bertan bilatuko da birtualizazio-plataformak euskarria eman ahal diezaien OAuth, OpenID, Kerberos eta SAML protokoloak.
- H. Faktore bikoitzeko autentifikatze-euskarria.** Plataformak euskarria eman ahal izan beharko dio faktore bikoitzeko autentifikatzeari, SPRIk zehaztu ahal dituen segurtasun-arauak betetze aldera.
- I. Erabiltzaile eta/edo roletan oinarritutako baimenetarako euskarria (RBAC).**

- J. Zehaztasun-maila handiko sarbide-kontrola.** Plataformaren bidez, SPRIk helburu izango du datuak zehaztasun-maila handiarekin murriztea, sarbide horiek errenkada eta zutabea murriztu ahal izatera heldu arte.

Birtualizazio-plataformak erabiltzailerako baimen zorrotz eta xehakatuak eta roletan oinarrituak ezarri ahalko ditu bertan zehaztutako elementu guztietarako. Baimenerako politikek zutabe eta errenkaden segurtasuna ezarri ahalko dute. Politika horiek honako konbinazio hauetako bat erabiliz zehaztu ahalko dira:

- Errenkada (hautaketa-baldintza bat aplikatuz).
- Zutabea.
- Errenkada eta Zutabea (errenkadei hautaketa-baldintza bat aplikatuz).

Baimen horiek kontsulta bakoitza egin aurretik ebaluatu beharko dira, erabiltzaileak emaitzak ikusteko baimenak dituen edo ez zehazteko. Baimenduta ez dauden datuak iragazi (hau da, kontsultako emaitzetatik kendu) edo maskaratu (hau da, existitzen direla egiaztatu, baina erabiltzaileak ezingo ditu ikusi) ahalko dira.

- K. Semantikan oinarritutako segurtasun-politika globalak.** Laburbilduz, SPRIk ahalegina egingo du plataforma gai izan dadin semantikan oinarritutako segurtasun-politikak kudeatzeko (adibidez, etiketak erabiliz). Adibidez, politika globala ezarri ahal beharko da, rol zehatza duten erabiltzaileek bakarrik izan dezaten “DIRULAGUNTZA” etiketa duten ikuspegiak kontsultatzeko baimena. Gainera, adibidez, etiketa zehatza duten zutabe guztietarako maskaratzea erabiltzen duen politika globala zehaztu ahal izan beharko da (“KONFIDENTZIALA”, “PRIBATUA” edo antzekoa).

## 2.4. OPTIMIZAZIOA ETA ERRENDIMENDUA

SPRIk antolakuntza-mailan datuen eremu zentralizatua eskuragarri jartzea nahi duen heinean, helburu du eremu hori osatzeko zein baliabideak ustiatzeko ahalik eta optimizatuen ere egon dadin, eragin ahalik eta txikiena izateko sistema operatiboetan, mantentze-lanetan, errendimendu baxutik eratorritako gertakarien kopuruan, erabiltzaileek esperientzia ona izan dezaten, etab.

Horretarako, datuak birtualizatzeko plataformari zenbait betekizun eskatuko zaizkio, eskakizun horiek betetze aldera.

- A. Kontsultak optimizatzea.** Plataformak kontsultak optimizatzeko teknikak eskaini beharko ditu errendimendua hobetzeko. Plataformak kontsultak datu-iturrietara eskuordetzea optimizatu beharko du, aurretik datu horiek birtualizazio-plataformara eraman edo bigarren gordailuan kopiatzeko beharrik egon gabe.

- B. Kostuetan oinarritutako optimizazioa.** Plataforma gai izan dadin kontsulta bakoitzerako egikaritze-plan hoberena hautatzeko, eta SPRIko langileek eskuz aztertu eta garatzeko beharrik gabe.
- C. AAn oinarritutako kontsultak bizkortzea.** Plataformak kontsultak bizkortzeko teknika adimendunak izan beharko ditu, AAn oinarrituak, etorkizunean SPRIn egon litezkeen analisi-egoera konplexuak ebazteko.
- D. Datuen zirkulazioa.** SPRIk uste du plataformak sareko zirkulazioa murriztu eta jatorrian kontsultak modu egokian egiteko dituen gaitasunak funtsezkoak direla. Birtualizazio-plataformak egikaritzea datuetan bertan egitea eta birtualizatzeko geruzan datuak nahastea murriztea (posiblea denean) lehenetsiko du. Gainera, teknika horiek aurretik nabarmendutako beste funtzioekin konbinatuz, hala nola informazioa cachean gordetzeko moduak, baliabideen erabilera egokia eta egungo datu operatiboen jatorrien gaineko erabilera ahalik eta txikiena lortu nahi dira.
- E. Egikaritze-planak.** Plataformak ad hoc kontsultak egitea ahalbidetu beharko du, eta, era berean, beharrezkoa da kontsulta horien egikaritzea egikaritze-plan dinamikoekin optimizatzea. Hala, eta normalean erabiltzaile teknikoek bakarrik dutenez optimizazioari buruzko oinarritzko ezagutza, funtzio horrek kontsultan ager litezkeen arazoak monitorizatu edo identifikatzeko balioko du.
- F. Arau estatikoak eta kostuetan oinarritutako optimizazioa.** Plataformak arau estatikoetan (heuristikoak) oinarritutako teknikak zein kostuetan oinarritutako optimizatzeko teknikak aplikatu beharko ditu. Gainera, plataformako kostuetan oinarritutako optimizagailuak automatikoki egikaritze-plan egokiena hauta dezan, zenbait faktore hartu beharko lirateke kontuan, hala nola:
- Estatistika-datuak, esate baterako, taulen tamaina, eremuen kopurua, eremu bakoitzaren tamaina, balio minimo eta maximoak, egon daitezkeen balio ezberdinak, etab.
  - Datuen iturriaren gaitasun espezifikoak, hala nola datuen iturri-mota (datu-basea, artxiboa, dokumentuak, etab.), modua (irakurtzeko bakarrik diren iturriak edo gorde daitezkeen datu-kopuru murriztua) edo tamaina (adibidez, MPP iturri bateko nodoen kopurua).
- G. Analisi-egoeretan kontsultak bizkortzea.** Plataformak datu gehitu eta summary-etarako euskarria izan beharko du, eta MPPrekin (Multi Parallel Processing) integratuta egon beharko da, datu-bolumen handien gaineko kontsultak egin ahal izateko.

- H. Gehikuntza-operazioak eskuordetzea.** Plataformak “group by” operazioak “join” edo “union” operazioen azpitik kokatu beharko ditu, berreskuratu beharreko erregistro-kopuruak murriztu ahal izateko.
- I. Partizionatuako loturak (Partition Pruning).** Lotura-ikuspegi bateko datuak datuen iturri bat baino gehiagotan banatzen denean, eta kontsultak lotura-iturriko datuak bakarrik behar dituztenean, birtualizazio-plataformako optimizagailuak kontsultan parte hartuko ez duten iturriak kendu beharko ditu.
- J. Datuak mugitzeko teknikak.** Teknika hori datu-multzo txiki bat beste iturri batean datu-multzo handiarekin konbinatu behar denean erabiliko da. Plataformako optimizagailua gai izan beharko da automatikoki datuak datu-multzo txikiena duen iturritik datu-multzo handiena duen iturri batera mugitzeko, aldi baterako taula batean. Hala, kontsulta osoaren egikaritzea gordailu bakar bati eskuordetu ahalko zaio, errendimendua nabarmenki hobetuz.
- K. Kontsulten bizkortze adimentsua.** Optimizatzailerak errendimenduan hobekuntza nabarmenak lortu beharko ditu, aurretik plataforman modu adimentsuan transakzio-datuen laburpenak (summary-ak, edozein prozesuan edo uneko prozesurako lortu/sortutako kontsulten emaitza) erabiliz, etorkizuneko kontsultak optimizatu ahal izateko.
- L. Laburpenak.** Administrazioak laburpenak sortu ahal beharko ditu, plataformak modu automatikoan egin ahal dizkion iradokizunak erabiliz. Emandako egoera bateko kontsulta-patroiak eta lehenagoko kontsulten errendimendua aztertzean, sistemak Machine Learning-eko ereduak aplikatuko du laburpen egokiak identifikatzeko (etorkizuneko kontsultak bizkortzen lagunduko dutenak). Laburpenak cachean edo kanpoko datu-iturri batean gorde ahalko dira.
- M. Monitorizazio-datuak.** Monitorizazio-datuak erabiliko dira (hala nola erabilera-metrikak, egikaritze-aldiak, datuen banaketa eta bolumenak, etab.) gomendioen motorra elikatzeke. Motor hori administrazioari errendimendua doitzeko laguntzeko erabili beharko da. Adimen artifiziala erabiliko da aurreko kontsultak aztertu eta kontsultak bizkortzeko “summary” berriak gomendatzeko.
- N. Prozesatze paraleloa.** Birtualizazio-plataformak prozesatze masibo paralelodun (MPP) motorrak dituen integratze natiboa izan beharko du, Big Data-ko egoera konplexuetan kontsultak bizkortze aldera. Plataformak operazioen prozesatze paraleloa egiteko gai izan beharko litzateke, beharrezkoa den guztietan prozesatze paraleloa erabiltzen duen kluster bati eskuordetuz, esate baterako, Spark, Databricks, Impala edo Presto bezalako teknologietan oinarritutako klusterrak.

Gainera, beharrezkoa da cachean gordetzeko aukera izatea, aipatutako teknologien barruan (Spark, Impala, Presto, etab.) duela gutxi erabilitako datuetarako sarbide azkarra izateko.

- O. MPP integratua.** Kotsuak bizkortzeko MPP mekanismoa integratuta egon behar da optimizatzeko ereduak, konponbidearen kostuetan oinarrituz. Plataformak azken erabakia hartuko du, azken prozesatzea memorian egiteko edo prozesatze hori klusterrari eskuordetzeko.

## 2.5. CLOUD

SPRIk hodeian aplikazioen eta produktuen arkitektura-konponbideak ditu. Hortaz, birtualizazio-plataformak euskarri oso eta integratua izan beharko du hodeiko iturri natiboetarako (datu-iturriak, cacheak edo erreplikazio selektiboko helmugak).

Konponbide hibridoak hedatzeen parte edo “hodeirako mugimenduen” planen katalizatzaile gisa, birtualizazio-plataforma bateragarria izan beharko da eta integratu ahal beharko da hodeiko hornitzaile nagusiekin, zehazki, AWS, Azure eta Google Cloud Platform. Hodeiko eremuetako berezko datu-iturriak, askotan SaaS modalitatean eskaintzen direnak, birtualizatzeko tresnak erabili ahal behar izango ditu hainbat zereginetarako: datu eskalagarrien iturri gisa, konponbidearen cache-sistema gisa edo materializatze selektiboaren helmuga gisa (aldi baterako taulak, summy-ak, etab.).

## 2.6. LOTEKA PROZESATZEA

SPRIk aplikazio, produktu eta kanpoko iturrien bidez bildutako informazio-bolumen handia du. Kasu askotan, informazio horren prozesatzea modu asinkronoan egin daiteke. Hortaz, birtualizazio-plataformak ohiko zereginak (cachean gordetzea, erreplikatzeko selektiboa, iturrietako aldaketak monitorizatzea, etab.) modu planifikatuan egiteko aukerak eman beharko ditu.

Plataformak zereginen programagailua eskaini beharko du batch zereginak egiteko. Honako hauek dira egon litezkeen ohiko erabilerak: datu-baseetara edo fitxategietara datuak esportatzea (fitxategi murriztuak, CSV, fitxategi lauak, MS Excel, aurkibideak, etab.), cachera kargatzea eta noizbehinkako antolaketa-zereginak (datuen estatistikak jasotzea, cachea eguneratzeko zereginen datuak, iturrietan aldaketak aldatzea, etab.) etab.

## 2.7. DEVOPS FUNTZIOAK

DevOps-ek garapen-taldeen eta IT zereginak egiten dituzten gainerako taldeen arteko elkarlana, integratzea, ikusgarritasuna eta gardentasuna sustatzen du. SPRIk DevOps azpiegitura du aplikazioak hedatzeko eta, hortaz, birtualizazio-plataformak sistema zentralizatua izan beharko du konponbidearen hedapenaren kudeaketa osoa egiteko, garapen-eremuetan eta ekoizpenean, on-premise eta hodeian zein eremu hibridoetan.

Gainera, birtualizazio-plataformak, derrigorrez, euskarria eman beharko du cloud hornitzaileetan hedapen automatizatuak egin ahal izateko, honako hauetan barne, baina ez bakarrik: AWS, Azure eta GCP. Horretaz gain, plataformak kudeaketa automatikoa eta monitorizatze integratua izan beharko du honako hauetarako: klusterrak, karga-orekagailuak, eremuen arteko migrazioa, autoeskalatzea, etab.

DevOps eremuaren barruan, plataformaren monitorizatzea denbora errealean eta osagai guztien gainean modu integratuan egin ahal beharko litzateke; gainera, plataforman egindako kontsulta, parametrizazio eta jarduera guztiak ikuskatu ahal beharko ditu.

Plataforman bilatzen diren beharretako beste batzuk dira hedatze konplexuen kudeatze zentralizatua egin ahal izatea (adibidez, kluster anitzak eta eskuragarritasun altuarekin) eta eguneratze, lizentzia-kode, eremu ezberdinen parametro espezifikoek (adibidez, datuen iturrietara konektatzeko parametroak), monitorizatzeen, log erregistroen eta auditoretzen kudeatze zentralizatua egitea.

## 2.8. HEDATZEAK

**A. Azpiegiturak.** Birtualizazio-plataformak euskarria eman beharko die honako azpiegitura hauetan hedatzeari:

- a. On-premise azpiegiturak.
- b. Hodei publiko eta pribatuko azpiegiturako eta hyperscaler nagusietan, hala nola AWS, Azure edo GCP.
- c. Azpiegitura hibridoak (on-premise + hodeia).

**B. Banatutako hedapenak.** Plataformak euskarria eman beharko du geografikoki banatutako hedapenetarako (errendimenduagatik edo araudia betetzeko arrazoiengatik).

**C. Ontziak.** Plataformak hedatu ahal izateko aukera eskaini beharko du, ontziak erabiliz (adibidez, Docker ontziak erabiliz).

## 2.9. LIZENTZIAMENDUA

**A. Plataforma.**

- B.** SPRIk hodeiko eta on-premise konponbideak dituenek, lizentzia hedapen-motatik berezita egon beharko da (on-premise, hodeia edo hibridoa), eta ezingo da lizentziarik aldatu edo kostu gehigarriak leporatuko plataforma aldatzeagatik.
- C. Dimentsionatzea/Kostuak.** SPRIk, eta Garapen Ekonomiko sailaren menpe dagoen eskubide pribatuko erakunde publikoa denez, zerbitzu publikoa ematen du, eta xede du datuak birtualizatzeko datua murriztuta ez egotea edo konektatutako kontsumitzaileen kopuruaren, datuen iturrietarako konexioen kopuruaren (adib., iturrien kopurua), edo birtualizatzeko motorrak prozesatutako datuen bolumenaren ondoriozko kostuak ez sortzea.

### **3. BESTE ESPEZIFIKAZIO BATZUK**

Honako hauek dira SPRIri datuak birtualizatzeko plataformari dagokionez balio erantsi handia ekarriko lioketen beste espezifikazio batzuk:

- A. Konponbide eskalagarria (horizontalki, bertikalki).
- B. Eskuragarritasun altua.
- C. Segurtasun-kopien eta ezbeharren aurrean berreskuratzeko sistemak.
- D. Metadatuak enkriptatzea.
- E. UI rol ezberdinetarako (administrariak, garatzaileak, datuen administrariak, erabiltzaile komertzialak, etab.).
- F. Konponbide integratua.
- G. Euskarria hizkuntza bat baino gehiagori.
- H. Estandar irekiak erabiltzearen bidez integratzea.
- I. MPP motor integratua: Hornitzaileak birtualizatzeko motorraren euskarriaren lizentzia sartuta dagoen motor horretarako euskarria eman beharko du.