



# **Plataforma transversal de dades**

## **Document d'arquitectura**

Responsable del document: Paul Bourne (arquitecte)  
Setembre 2024

# Índex

1.	Introducció	3
1.1	Propòsit	3
1.2	Abast	3
1.2.1	Necessitats fonamentals	3
1.3	Parts interessades	4
2.	Vistes	5
2.1	Vista de Context	5
2.2	Vista Funcional	10
2.3	Vista d'Informació	13
2.4	Vista de Concurrencia	14
2.5	Vista de Desenvolupament	15
2.6	Vista de Desplegament	16
2.6.1	Prerequisit per a la creació de l'entorn	16
2.6.2	MongoDB Atlas	17
2.6.3	Azure Databricks i Unity Catalog	20
2.6.4	Denodo	23
2.6.5	Altres recursos a desplegar	24
2.6.6	Altres dades rellevants pel desplegament	25
2.6.7	Vista Operacional	25
3.	Perspectives Transversals	27
3.1	Seguretat	27
3.2	Rendiment i escalabilitat	29
3.3	Disponibilitat	29
3.4	Localització	30
4.	Informació específica pel projecte d'aprovisionament	31
4.1	Informació relativa al context	31
4.2	Informació relativa al SIC	31
4.3	Informació relativa a xarxes i dominis DNS	31
4.4	Informació relativa a l'aprovisionament d'Infraestructura	39
5.	Referències	40

## 1. Introducció

### 1.1 Propòsit

L'objectiu de l'arquitectura es posar en marxa una Plataforma Transversal de Dades per als serveis de dades de la Generalitat de Catalunya i establir criteris, polítiques i pautes de caràcter transversal per a les diferents necessitats de gestió i explotació de dades.

### 1.2 Abast

Disposar d'una plataforma que permeti agregar dades de diverses fonts facilitant una visió unificada d'elles, tant en temps real com en batch, amb paràmetres d'alta qualitat i confiabilitat, publicades en un catàleg central i fàcilment accessibles per al seu consum autònom, amb diferents nivells d'accés, visibilitat i seguretat. A més a més, servirà de base per a realitzar els diferents nivells d'anàlisi de les dades que posi a disposició dels usuaris.

#### 1.2.1 Necessitats fonamentals

##### Requisits subscripcions:

- Es requereix d'una subscripció a Microsoft Azure, AWS i Google Cloud Platform.
- Es requereix d'una subscripció a Databricks per cadascun dels clouds prèviament esmentades.
- Es requereix d'una subscripció a MongoDB Atlas en Azure.
- Les regions dels clouds i de tots els components han de pertànyer a la Unió Europea.
- Es requereixen permisos d'Administració sobre les subscripcions cloud (Microsoft Azure, AWS i Google Cloud Platform):
  - Rol de propietari de la subscripció.
  - Rol de desenvolupador d'aplicacions o superior en el Active Directory.

Veure requisits amb més detall a la documentació de Databricks [2] i MongoDB Atlas [1].

##### Restriccions i requisits no funcionals

Principals requisits de recursos i serveis necessaris per a la infraestructura:

- Requisits i restriccions de Databricks, veure [2].
- Requisits i restriccions de MongoDB Atlas, veure [1].

### 1.2.1.2 Restriccions i requisits no funcionals de Microsoft Azure

- Private Links entre les VNets de Databricks i MongoDB Atlas.
- Azure Key Vault.
- **Azure ExpressRoute** entre les instal·lacions de CTTI i Azure.
- Azure Active Directory
- ADLS Gen2

## 1.3 Parts interessades

S'han identificat les següents parts interessades en la solució:

- Promotor:
  - CTTI / DGAD
  - Gestor / Responsable de la solució: Joan Duran Bofarull
- Responsable del projecte: Gestor d'Integració de Solucions: Joan Duran Bofarull
- Equip de desenvolupament: Minsait
- Equip de proves: Minsait, CTTI, departaments
- Equip de projectes d'Infraestructura: CTTI, Cloud i MongoDB Atlas.
- Equip d'administració i explotació de sistemes: CTTI, Cloud i MongoDB Atlas.
- Equip d'administració i explotació de xarxes: Nus
- Oficina de Seguretat: Agència de Ciberseguretat
- Equip de Qualitat: Oficina de Qualitat CTTI
- Equip d'administració i operacions funcionals
- Aplicació: **Codi 3623**- Plataforma Transversal de Dades del CTTI

## 2. Vistes

### 2.1 Vista de Context

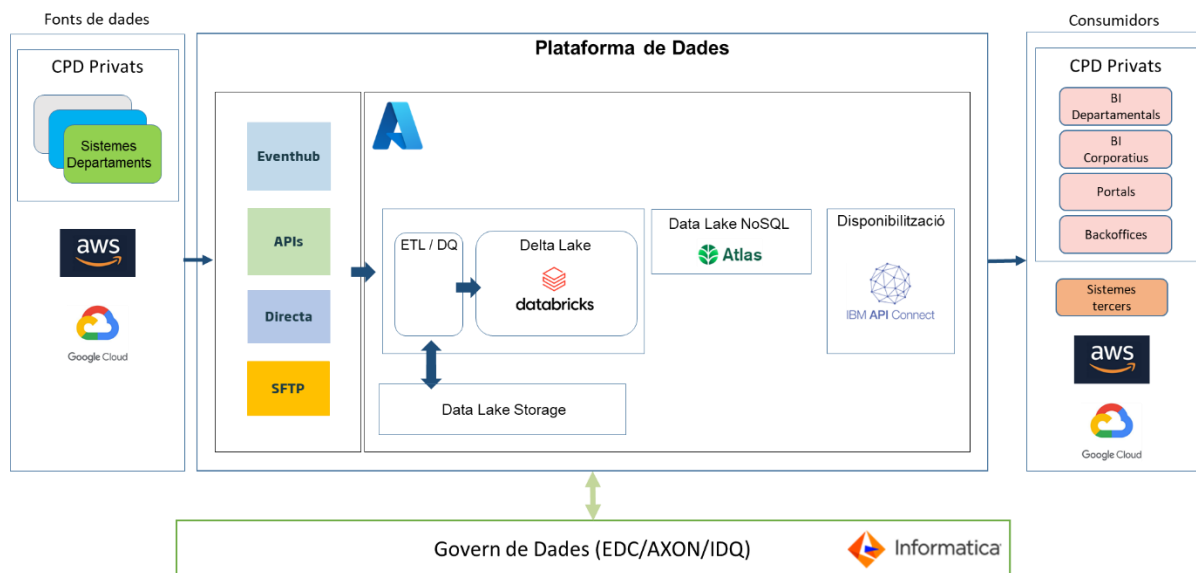
Els sistemes del CTTI i departaments de la Generalitat de Catalunya podran intercanviar informació amb la Plataforma Transversal de Dades, com es pot veure al diagrama de context de la Figura 1. La Plataforma Transversal de Dades intercanviarà informació amb *Informatica Axon* per poder dur a terme el govern de les dades de la plataforma.

La plataforma serà multicloud (Azure, AWS, GCP), però en una primera iteració s'implantarà només sobre Azure.

**Entitats externes a nivell funcional:**



### Diagrama de Context a nivell funcional



• **Figura 1 – Diagrama de context de l'arquitectura**

Descripció dels diferents sistemes externs utilitzats:

Sistema	Descripció	Localització (CPD)
Eventhub	Plataforma Kafka del CTTI. Encara que formalment és part de la Plataforma Transversal de Dades, serà un dels principals sistemes emprats per a la ingestió de dades dins la mateixa. Les aplicacions dels departaments publicaran esdeveniments (amb les dades a <i>ingestar</i> ), que seran consumits i tractats pel mòdul ETL/DQ de la plataforma.	CPD4
EDC / Axon / IDQ	Plataforma de govern i qualitat de les dades del CTTI. Obtindrà i actualitzarà metadades de la plataforma.	Cloud Azure
Sistemes Departaments	Sistemes i aplicacions del CTTI o d'altres departaments de la Generalitat que es troben dins els seus <b>CPDs</b> . Seran els encarregats de publicar informació a la Plataforma Transversal de Dades. En alguns casos, també podran consumir dades de la plataforma.	Dependrà del sistema.
BI Departamentals / Corporatius	Sistemes de Business Intelligence, tant dels diferents departaments com a corporatius del CTTI.	Dependrà del BI.
Portals	Portals que puguin fer servir els diferents departaments per a recollir informació de la Plataforma Transversal de Dades.	Dependrà del portal.
Backoffices	Backoffices que puguin implantar els diferents departaments per a recollir informació de la Plataforma Transversal de Dades.	Dependrà del backoffice.
Sistemes tercers	Sistemes de tercers o fora de l'àmbit de CTTI.	Dependrà del sistema.
Altres Clouds	Interacció amb altres clouds, com Amazon AWS i Google Cloud Platform.	N/A

Detall de la interacció entre el sistema / solució i els sistemes externs:

Origen	Destí	Informació intercanviada	Característiques interacció	
Sistemes Departaments	Plataforma Transversal de Dades	Esdeveniments	Estil integració (Principi arquitectura 1.3)	<input type="checkbox"/> Web Service <input type="checkbox"/> Api <input type="checkbox"/> Sftp <input type="checkbox"/> Rpc <input type="checkbox"/> Cues <input type="checkbox"/> Altres: tòpics (esdeveniments)
			Temporalitat o periodicitat	Cada cop que es produeix un esdeveniment nou.
			Consideracions	Actualment existeix ja aquesta integració.

Origen	Destí	Informació intercanviada	Característiques interacció	
				Cada nova aplicació que s'hi vulgui integrar, ha de realitzar una petició a l'Oficina Tècnica d'Eventhub.
Plataforma Transversal de Dades	EDC / Axon / IDQ	Metadades	Estil integració (Principi arquitectura 1.3)	<input type="checkbox"/> Web Service <input type="checkbox"/> Api <input type="checkbox"/> Sftp <input type="checkbox"/> Rpc <input type="checkbox"/> Cues <input type="checkbox"/> Altres: HTTPS, JDBC
			Temporalitat o periodicitat	
			Consideracions	EDC / Axon s'encarrega de fer peticions d'extracció de metadades sobre la Plataforma Transversal de Dades.
Sistemes Departaments	Plataforma Transversal de Dades	Arxius de dades i documents a ingestar en la plataforma.	Estil integració (Principi arquitectura 1.3)	<input type="checkbox"/> Web Service <input type="checkbox"/> Api <input type="checkbox"/> Sftp <input type="checkbox"/> Rpc <input type="checkbox"/> Cues <input type="checkbox"/> Altres
			Temporalitat o periodicitat	
			Consideracions	La PTD podrà fer servir SFTP per tal d'ingestar dades. Ha de poder accedir a l'SFTP.
Plataforma Transversal de Dades	BI	Dades de les taules de Delta, col·leccions MongoDB Atlas i taules PostgreSQL.	Estil integració (Principi arquitectura 1.3)	<input type="checkbox"/> Web Service <input type="checkbox"/> Api <input type="checkbox"/> Sftp <input type="checkbox"/> Rpc <input type="checkbox"/> Cues <input type="checkbox"/> Altres: JDBC, Denodo, Delta Share
			Temporalitat o periodicitat	
			Consideracions	Els BI realitzarà <i>queries</i> contra la Plataforma Transversal de Dades.
Plataforma Transversal de Dades	Portals dels sistemes d'informació	Dades ADLS, de les taules de Delta, col·leccions MongoDB Atlas	Estil integració (Principi arquitectura 1.3)	<input type="checkbox"/> Web Service <input type="checkbox"/> Api <input type="checkbox"/> Sftp <input type="checkbox"/> Rpc <input type="checkbox"/> Cues

Origen	Destí	Informació intercanviada	Característiques interacció	
		i taules PostgreSQL.		<input type="checkbox"/> Altres: JDBC, Denodo, Delta Share, SMB/NFS
			Temporalitat o periodicitat	
			Consideracions	Els portals realitzaran peticions contra la Plataforma Transversal de Dades.
Plataforma Transversal de Dades	Backoffices	Dades ADLS, de les taules de Delta, col·leccions MongoDB Atlas i taules PostgreSQL.	Estil integració (Principi arquitectura 1.3)	<input type="checkbox"/> Web Service <input type="checkbox"/> Api <input type="checkbox"/> Sftp <input type="checkbox"/> Rpc <input type="checkbox"/> Cues <input type="checkbox"/> Altres: JDBC, Denodo, Delta Share, SMB/NFS, tòpics (esdeveniments)
			Temporalitat o periodicitat	
			Consideracions	Els Backoffices realitzaran peticions contra la Plataforma Transversal de Dades.
Plataforma Transversal de Dades	Sistemes de tercers	Dades ADLS, de les taules de Delta, col·leccions MongoDB Atlas i taules PostgreSQL.	Estil integració (Principi arquitectura 1.3)	<input type="checkbox"/> Web Service <input type="checkbox"/> Api <input type="checkbox"/> Sftp <input type="checkbox"/> Rpc <input type="checkbox"/> Cues <input type="checkbox"/> Altres: Delta Share, Denodo
			Temporalitat o periodicitat	
			Consideracions	Els sistemes de tercers realitzaran peticions contra la Plataforma Transversal de Dades.
Gestor d'APIs del CTTI	Plataforma Transversal de Dades	Dades ADLS, de les taules de Delta, col·leccions MongoDB Atlas i taules PostgreSQL.	Estil integració (Principi arquitectura 1.3)	<input type="checkbox"/> Web Service <input type="checkbox"/> Api (REST) <input type="checkbox"/> Sftp <input type="checkbox"/> Rpc <input type="checkbox"/> Cues <input type="checkbox"/> Altres: JDBC
			Temporalitat o periodicitat	
			Consideracions	Possibilitat de treballar amb WS. Preferiblement anar cap a API REST.
Altres Clouds	Plataforma Transversal de Dades	Arxius de dades i documents a	Estil integració (Principi arquitectura 1.3)	<input type="checkbox"/> Web Service <input type="checkbox"/> Api (REST) <input type="checkbox"/> Sftp

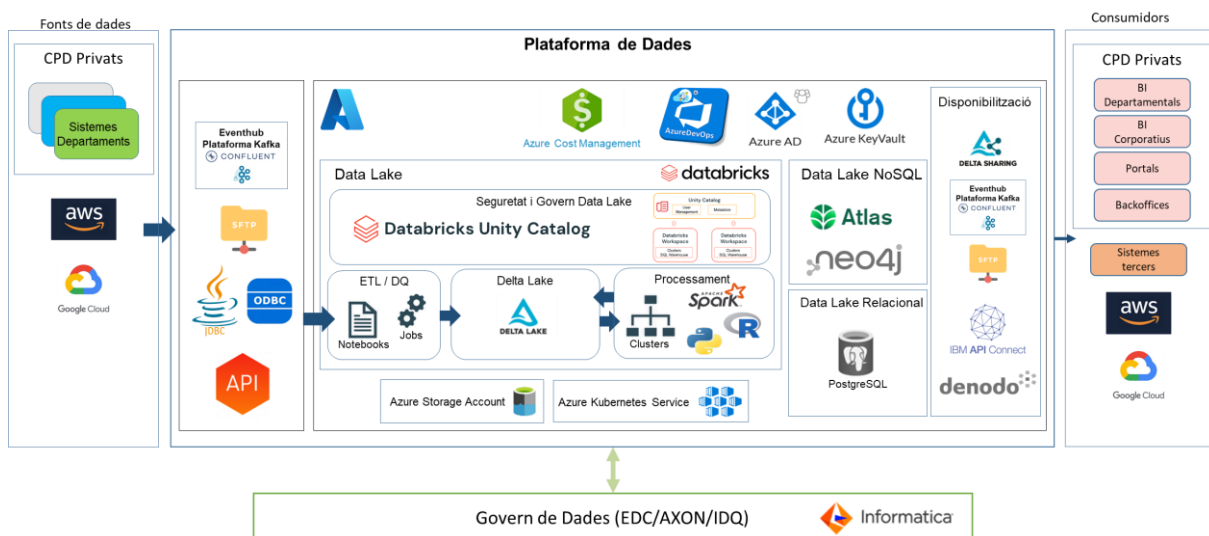


Origen	Destí	Informació intercanviada	Característiques interacció	
		ingestar en la plataforma.		<input type="checkbox"/> Rpc <input type="checkbox"/> Cues <input type="checkbox"/> Altres: JDBC, Denodo, tòpics (esdeveniments)
			Temporalitat o periodicitat	
			Consideracions	
Plataforma Transversal de Dades	Altres Clouds	Arxius de dades i documents a ingestar en la plataforma.	Estil integració (Principi arquitectura 1.3)	<input type="checkbox"/> Web Service <input type="checkbox"/> Api (REST) <input type="checkbox"/> Sftp <input type="checkbox"/> Rpc <input type="checkbox"/> Cues <input type="checkbox"/> Altres: JDBC, Denodo, tòpics (esdeveniments)
			Temporalitat o periodicitat	
			Consideracions	

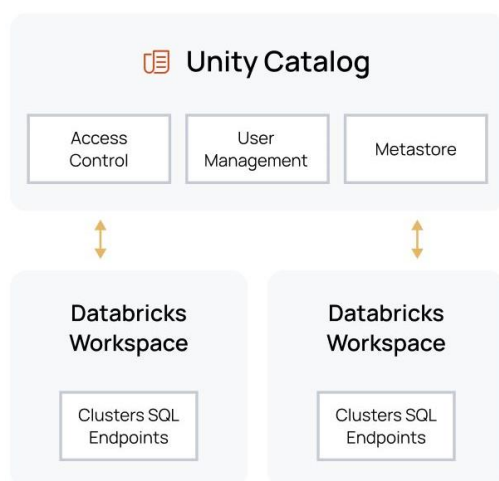
### Perfils d'Actors:

Perfil Actor	Descripció
Aplicació	Escriptura i lectura d'informació al Data Lake (Delta, ADLS, MongoDB Atlas, PostgreSQL, Neo4j, etc.).
Càrrega	Gestió de càrrega: extracció, transformació i càrrega a la plataforma.
Explotació	Explotació, transformació i processament de dades.
Lectura	Visualització de dades tractades a la Plataforma. Aquests usuaris poden ser d'aplicacions externes que consumeixen la informació de la Plataforma Transversal de Dades (ODI, OBIEE, BI, Notebooks, portals, etc.).
Govern	Control de les metadades i compliment de polítiques de la Plataforma.
Data Scientist	Creació i gestió de models d'analítica avançada.
Administradors	Administració de les diferents plataformes que conformen la Plataforma Transversal de Dades: Azure de CTTI, Databricks, MongoDB Atlas i Eventhub.
Departaments	Usuaris dels departaments que han de treballar amb les dades de referència de la plataforma.

## 2.2 Vista Funcional



• **Figura 2 – Diagrama funcional de l'arquitectura**



Mitjançant la funcionalitat Databricks Workspace, que proveix Unity Catalog, els departaments tenen la possibilitat de mantenir els seus processos i dades aïllats de la resta de departaments, sense afectar a nivell de recursos a la resta d'execucions.

Les dades també estaran separades i cada departament veurà per defecte només les seves, excepte quan explícitament es decideix compartir-les.

• **Figura 3 – Esquema de Unity Catalog**

**Estructura de mòduls funcionals interna del sistema:**

Mòdul funcional	Descripció
Ingesta	El mòdul d'ingesta permet la ingestió de dades des dels sistemes fonts. En aquesta fase estarà format per la plataforma <b>Eventhub</b> (Plataforma Kafka del CTTI), desplegada a les instal·lacions de CTTI, SFTP,...

Mòdul funcional	Descripció
Eventhub	<p>Plataforma Kafka del CTTI. Un dels principals sistemes emprats per a la ingestió de dades dins la Plataforma Transversal de Dades. Les aplicacions dels departaments publicaran esdeveniments (amb les dades a ingestar), que seran consumits i tractats pel mòdul Flow Management de la plataforma.</p> <p>La comunicació entre Eventhub i PTD hauria de passar a través d'una connexió <b>ExpressRoute</b> que permeti l'intercanvi d'informació entre <b>Eventhub, la NET0 i la VNET de la PTD</b>.</p>
SFTP	La PTD s'ha d'integrar amb el servei SFTP corporatiu del CTTI, a on els seus clients podran deixar arxius a ingestar. Aquests fitxers es copiaran dins l'emmagatzemament Azure de PTD, que seran recollits i tractats pel seu mòdul d'ingesta.
JDBC/ODBC	La PTD s'ha d'integrar a orígens de bases de dades mitjançant JDBC/ODBC
API	La PTD s'ha d'integrar a fonts d'informació publicades amb API Rest o WS.
Data Lake NoSQL	Aquest mòdul estarà implementat sobre MongoDB Atlas i neo4j (AuraDB), que permeten el tractament de BBDD NoSQL en <i>cloud</i> . El <i>cloud</i> seleccionat de MongoDB Atlas serà el de Microsoft Azure.
Azure Active Directory	Serà el corporatiu del CTTI.
Azure Storage Account	Comptes d'emmagatzematge d'Azure que contenen els objectes de dades necessaris per estovar les dades de la PTD i MongoDB Atlas, que els gestionaran. S'hauran de configurar perquè emprin les claus de xifrat en repòs guardades a Azure Key Vault, gestionades per CTTI.
Azure Key Vault	Permet administrar les claus criptogràfiques de les plataformes i serveis desplegats sobre el cloud de Microsoft Azure. Les claus criptogràfiques hauran de ser gestionades, custodiades, arxivades i destruïdes en tot moment pel CTTI, només per usuaris autoritzats per aquest; mai per Azure ni MongoDB Atlas.
Azure DevOps	Proporciona un conjunt integrat de serveis i eines per gestionar els vostres projectes de programari, des de la planificació i el desenvolupament a través de proves i desplegament.
Azure Cost Management	Proporciona informació sobre els costos generals i la utilització en tots els serveis d'Azure.
Databricks	Plataforma d'anàlisi oberta i unificada per construir, desplegar, compartir i mantenir dades, anàlisis i IA a gran escala.
Seguretat i Govern Data Lake	Format per Unity Catalog que proporciona capacitats de control d'accés centralitzat, auditoria, llinatge i descobriment de dades a través dels espais de treball d'Azure Databricks.

Mòdul funcional	Descripció
ETL/DQ	Conjunt d'eines destinades a l'extracció, càrrega i transformació de dades. De la mateixa manera es disposa de DQ que comprova la qualitat d'aquestes.
Delta Lake	Capa d'emmagatzematge que proporciona la base per a l'emmagatzematge de dades i taules a la Plataforma Databricks.
Processament	Mòdul encarregat de la transformació de les dades.
Azure Services      Kubernetes	Component que facilita la implementació, administració i escalat d'aplicacions en contenidors. AKS s'encarrega de tasques complexes com la configuració i el manteniment del clúster, permetent enfocar-te en el desenvolupament i l'operació de les aplicacions.
Data Lake Relacional	Base de dades relacional: PostgreSQL, Azure SQL, etc.
Disponibilització	La disponibilització de les dades es farà per diverses opcions: Delta Share, Eventhub, SFTP, IBM API Connect i Denodo.

#### Serveix Externs:

Serveis externs	Opcions
<b>SGDE</b>	<input type="checkbox"/> Formularis PDF <input type="checkbox"/> Formularis HTML5 <input type="checkbox"/> STD (Transf. documents) <input type="checkbox"/> No
<b>GECO+</b>	<input type="checkbox"/> Si <input type="checkbox"/> No
<b>PICA</b>	<input type="checkbox"/> Si <input type="checkbox"/> No
<b>Altres Integracions</b>	

#### Justificacions de les decisions del model funcional

Decisió	Justificació (Avantatges i Inconvenients)
Segregació funcional dels mòduls	Aplicació del principi d'única responsabilitat per els components de la plataforma.

## 2.3 Vista d'Informació

Dada a proporcionar	Opcions
<b>Dades de caràcter personal</b>	<input type="checkbox"/> Sí <input type="checkbox"/> No
<b>Finalitat i ús de les dades (RGPD)</b>	<p>S'estan definint els màxims nivells de seguretat de les dades, perquè preveiem que es puguin tractar i consolidar tots els nivells definits a la RGPD. L'objectiu és centralitzar l'emmagatzematge i gestió d'informació a ser utilitzada per diferents sistemes. No es posen límits al grau de sensibilitat que pugui allotjar.</p>
<b>Nivell de RGPD assignat al fitxer</b>	<input type="checkbox"/> Dades Bàsiques <input type="checkbox"/> Especialment protegides
<b>Nivell de sensibilitat de les dades</b>	<input type="checkbox"/> Públic <input type="checkbox"/> Intern <input type="checkbox"/> Sensible <input type="checkbox"/> Crític <input type="checkbox"/> Molt Crític
<b>Requeriment legal de retenció de les dades</b>	<input type="checkbox"/> 1 any <input type="checkbox"/> 2 anys <input type="checkbox"/> 3 anys <input type="checkbox"/> 4 anys <input type="checkbox"/> 5 anys: dades financeres <input type="checkbox"/> Altres: el que marqui la llei en cada cas.
<b>Model d'emmagatzematge de la Informació</b>	<input type="checkbox"/> Operacional <input type="checkbox"/> Analítiques <input type="checkbox"/> Documentals <input type="checkbox"/> Textuals <input type="checkbox"/> Cache <p>Totes les dades seran xifrades en repòs, emprant les tècniques d'Azure i amb les claus de xifratge gestionades per CTTI a Azure Key Vault.</p>
<b>Volumetries esperades d'informació</b>	<p>No tenim especificat unes volumetries perquè s'estan estudiant els casos d'ús. El dimensionament definit en la vista de desplegament es una línia base mínima per desplegar en el cloud. L'emmagatzemament i el processament són elàstics.</p>

## Entitats de referència:

Dada a proporcionar	Opcions									
Entitats de referència utilitzades	No aplica <table border="1"> <thead> <tr> <th>Grup</th><th>Entitat</th><th>Versió</th></tr> </thead> <tbody> <tr><td> </td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td> </td></tr> </tbody> </table>	Grup	Entitat	Versió						
Grup	Entitat	Versió								
Noves entitats a afegir	<input type="checkbox"/> Si, les següents: <table border="1"> <thead> <tr> <th>Grup</th><th>Entitat</th></tr> </thead> <tbody> <tr><td> </td><td> </td></tr> </tbody> </table> <input type="checkbox"/> No	Grup	Entitat							
Grup	Entitat									

## 2.4 Vista de Concurrencia

Dada a proporcionar	Opcions / Detall
Usuaris simultanis	La Plataforma Transversal de Dades albergarà diferents serveis que poden ser consumits per transaccions d'aplicacions, sistemes de BI, processos d'analítica. Per tant els usuaris simultanis poden contemplar-se com el total d'usuaris, transaccions o processos externs que consumiran serveis de la plataforma.
Identificació de processos	No s'identifiquen processos interns que generin pics de consum. En qualsevol cas, la plataforma és elàstica i s'ajustarà a les necessitats en cada moment.  Pel que fa als processos de govern, cal identificar possibles procediments de clients que poden afectar a la resta dels serveis.
Relació / comunicació entre processos	No hi ha col·lisió entre processos.

## 2.5 Vista de Desenvolupament

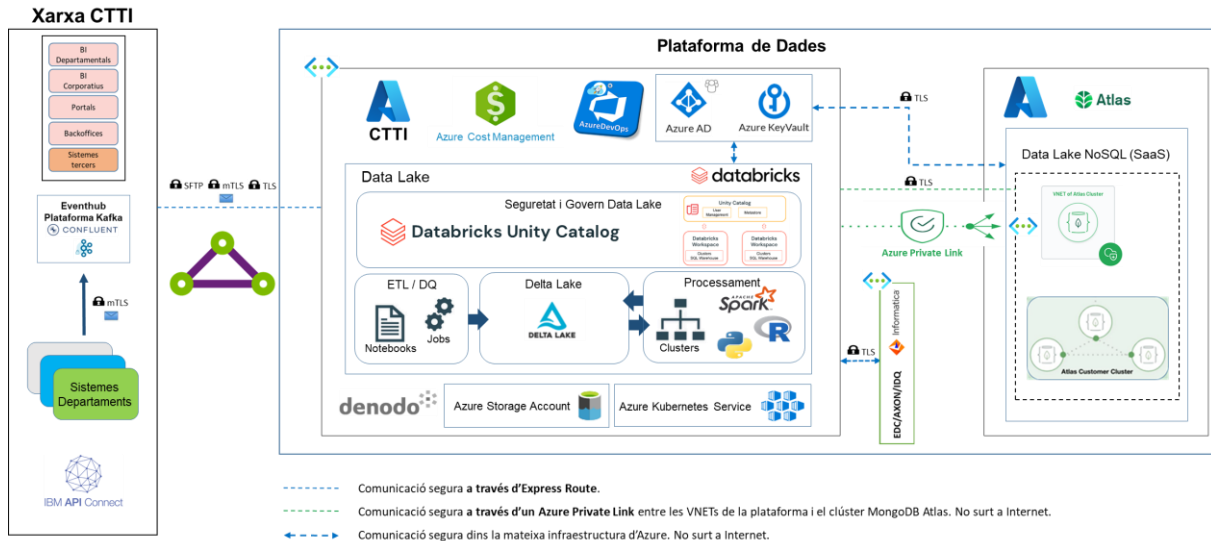
Dada a proporcionar	Opcions / Detall
<b>Tecnologies de desenvolupament</b>	Productes de Databricks (Spark, Delta), MongoDB Atlas, Confluent Kafka i llenguatges de programació com Scala, Python i Java.
<b>Identificar software / Llibreries de tercers utilitzades.</b>	Productes de Databricks (Spark, Delta), MongoDB Atlas, Confluent Kafka, entre d'altres.
<b>Principis i estàndards seguits en el disseny i desenvolupament del codi</b>	<p>Indicar els <a href="#">Principis d'Arquitectura</a> que es segueixen respecte al disseny i el desenvolupament.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Segregació de funcions</li> <li><input type="checkbox"/> Arquitectura desacoblada</li> <li><input type="checkbox"/> Arquitectura orientada a Serveis</li> <li><input type="checkbox"/> Reutilització de funcions</li> <li><input type="checkbox"/> Altres: Arquitectura d'esdeveniments</li> </ul>
<b>Repositori de codi</b>	<p>Informació del repositori on es puja el codi font.</p> <p>Repositoris generals:</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Git Corporatiu (GitLab – Per al codi que es desenvoluparà amb Scala, Python, Java i Terraform, entre d'altres)</li> <li><input type="checkbox"/> Host, SAP o paquets</li> </ul> <p>Repositoris particulars departamentals:</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Bamboo de Salut</li> <li><input type="checkbox"/> SVN d'Agaur</li> <li><input type="checkbox"/> SVN de l'Agència d'Habitatge</li> <li><input type="checkbox"/> SVN de TSF</li> <li><input type="checkbox"/> SVN d'Incasòl</li> <li><input type="checkbox"/> SVN d'Interior</li> <li><input type="checkbox"/> Quickbuild de Presidència</li> </ul> <p><input type="checkbox"/> Altres / Excepcions: Les parts propietàries de la plataforma no faran servir cap repositori.</p>
<b>Identificar jocs de caràcters</b>	<p>Indicar quin serà el joc de caràcters que s'utilitzarà.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> UTF8</li> <li><input type="checkbox"/> UTF16</li> <li><input type="checkbox"/> ISO8859-P15</li> <li><input type="checkbox"/> Altres:</li> </ul> <p>Nota: En un servei nou s'ha de fer ús d'UTF8</p>

## Justificacions de les decisions de la vista de desenvolupament

Decisió	Justificació
Databricks	Entorn d'execució distribuït i multicloud.
MongoDB Atlas	Compleix amb els requisits d'excel·lència tècnica i suport empresarial que es requereix per a una instal·lació transversal que donarà suport a sistemes crítics.
Confluent Kafka	Compleix amb els requisits d'excel·lència tècnica i suport empresarial que es requereix per a una instal·lació transversal que donarà suport a sistemes crítics

## 2.6 Vista de Desplegament

Es desplegarà la solució de Databricks allotjat al núvol de Microsoft Azure de CTTI mitjançant la filiació de la subscripció del CTTI en Azure amb una relació de confiança. També s'ha de desplegar la solució MongoDB Atlas en la modalitat **SaaS** de cloud públic allotjat al núvol de Microsoft Azure. Es realitza la subscripció directament amb MongoDB Atlas. A la Figura 4 s'observa l'esquema detallat de desplegament. Pel que fa a la relació entre els components, es pot veure a l'esquema de la Figura 2.



• **Figura 4 – Diagrama de desplegament de l'arquitectura**

### 2.6.1 Prerequisit per a la creació de l'entorn

Veure apartat 1.2.1.



## 2.6.2 MongoDB Atlas

El prerequisits per poder donar d'alta i treballar amb el servei SaaS MongoDB Atlas són els següents:

- Enviar a MongoDB el *tenant id* del compte d'Azure de la PTD.
- MongoDB retornarà un codi d'activació per tal de que sigui activat des de la seva plataforma.
- A MongoDB Atlas es crea una organització i dos projectes diferents, un per a PRE i l'altre per a PRO.

Encara que MongoDB Atlas és un SaaS, aquest obliga a que el client defineixi la següent informació mínima per a un *replica set* de 3 nodes:

- Proveïdor del cloud i regió: **Microsoft Azure** i *West Europe*.
- Tamany de la màquina sobre el que es desplegarà el servei.
- Versió de MongoDB.
- Backup i tipus de backup.

### Entorn Preproducció

#### Cloud Públic:

Preproducció - SaaS					
CONTENIDORS					
Identificador d'instància	Nombre Pods/ Contenidors	Programari i versió / Imatge Docker	Memòria Ram i Recursos addicionals	Disc Persistent	Administrat per CPD (Si/No)
Capa Aplicacions					
1	3	MongoDB 6	2 vCPUs – 8 GB	128 GB	No

A les dues imatges següents es pot veure la configuració requerida per a aquest entorn:

## Dedicated Clusters for high-traffic applications and large datasets

- Additional hardware configurations available for specialized workloads

	Tier	RAM	Storage	vCPU	Price
✓	M30	8 GB	32 GB	2 vCPUs	from \$0.68/hr
	Class	General			
	Storage	32 GB is included in the base price.			
		<div>8 GB<div></div>512 GB</div>			<div>128GB</div>
	Auto-scale	<div><input checked="" type="checkbox"/> Cluster Tier Scaling <a href="#">View docs</a></div> <div><div>Minimum cluster size<div>M30</div></div><div><input checked="" type="checkbox"/> Allow cluster to be scaled down</div></div> <div><div>Maximum cluster size<div>M40</div></div></div>			
		<div><input checked="" type="checkbox"/> Storage Scaling ⓘ</div>			
	IOPS	500 IOPS (48 MB/s throughput)			
	Additional Info	3000 max connections   High network performance			

### Additional Settings

#### MongoDB 6.0, Backup, Encryption at Rest

Cloud Backup

##### Select a Version

All clusters launch with the WiredTiger™ storage engine.

MongoDB 6.0

##### Turn on Cloud Backup (M2 and up)

Snapshots are taken automatically and stored according to your backup and retention policy. Meets recovery point objective (RPO) of 6 hours by default, configurable down to 1 hour.

You can disable backups at any time. [Learn more about backup options](#)

Pricing Varies

##### Continuous Cloud Backup

This additional option also records the full oplog for a configured window, permitting a restore to any [point in time](#) within that window. Meets recovery point objective (RPO) of 1 minute.

## Entorn Producció


### Cloud Públic:

Producció - SaaS					
CONTENIDORS					
Identificador d'instància	Nombre Pods/ Contenidors	Programari i versió / Imatge Docker	Memòria Ram i Recursos addicionals	Disc Persistent	Administrat per CPD (Si/No)
Capa Aplicacions					
2	3	MongoDB 6	2 vCPUs – 8 GB	512 GB	No


A les dues imatges següents es pot veure la configuració requerida per a aquest entorn:

#### Dedicated Clusters for high-traffic applications and large datasets

- Additional hardware configurations available for specialized workloads

Tier		RAM	Storage	vCPU	Price
✓	M30	8 GB	32 GB	2 vCPUs	from \$0.68/hr
	Class	General			
	Storage	32 GB is included in the base price.			
		8 GB  512 GB <div>512 GB</div>			
	Auto-scale	<input checked="" type="checkbox"/> Cluster Tier Scaling <a href="#">View docs</a> <div>             Minimum cluster size <span>M30</span>             Maximum cluster size <span>M40</span> </div> <input checked="" type="checkbox"/> Allow cluster to be scaled down			
	IOPS	2300 IOPS (48 MB/s throughput)			
Additional Info		3000 max connections   High network performance			

### Additional Settings


MongoDB 6.0, Backup, Encryption at Rest   
Cloud Backup

Select a Version

MongoDB 6.0

All clusters launch with the WiredTiger™ storage engine.

---

Turn on Cloud Backup (M2 and up) 


Snapshots are taken automatically and stored according to your backup and retention policy. Meets recovery point objective (RPO) of 6 hours by default, configurable down to 1 hour.

You can disable backups at any time. [Learn more about backup options](#)

Pricing Varies

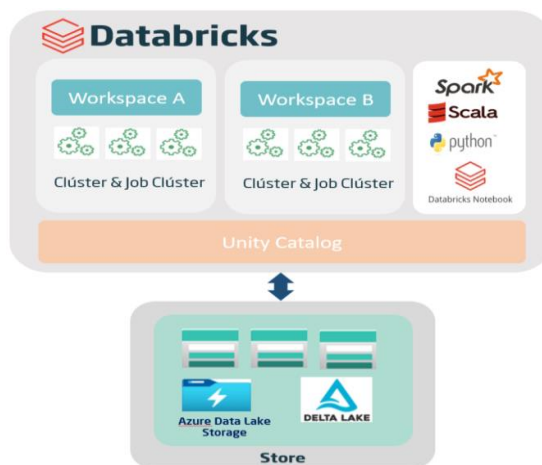
Continuous Cloud Backup

This additional option also records the full oplog for a configured window, permitting a restore to any [point in time](#) within that window. Meets recovery point objective (RPO) of 1 minute.

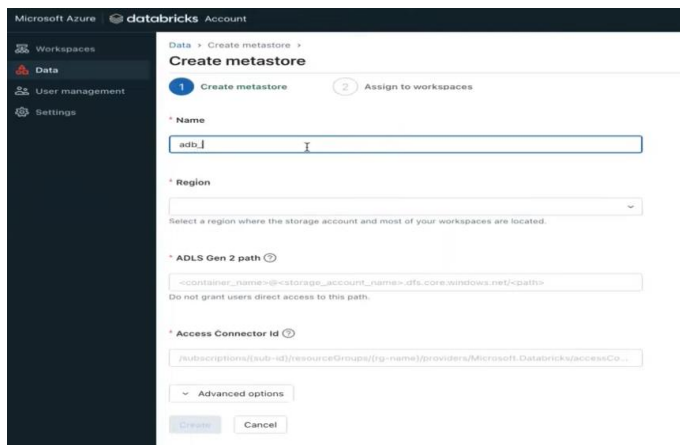


### 2.6.3 Azure Databricks i Unity Catalog

La implementació de Databricks es realitza eficientment a través de Terraform, aprofitant els mòduls especialitzats dissenyats per crear de manera integral tots els components essencials de la instància o workspace. Posteriorment, un cop aprovisionat Databricks, es procedeix a habilitar l'Unity Catalog per gestionar les dades emmagatzemades a la plataforma. Per dur a terme aquesta tasca, és fonamental comptar amb un usuari designat com a Global Admin del tenant al compte d'Azure.



L'habilitació de l'Unity Catalog implica la creació d'un metastore sobre un storage account d'Azure, tot recordant que només es pot crear un metastore per regió. Aquest Unity Catalog atorga una visió integral de tots els espais de treball creats sota el tenant, possibilitant un govern efectiu de la informació a través de fronteres i assegurant-ne una administració coherent i centralitzada. Els noms dels atributs seguiran les directrius del llibre blanc CTTI



- **Workspace:** És un contenidor que agrupa i organitza els recursos relacionats amb un projecte, aplicació o iniciativa. Proporciona un entorn unificat per a l'administració, col·laboració i desenvolupament de solucions al núvol.
- **Clúster:** Un clúster a Azure Databricks és un conjunt de màquines virtuals que treballen juntes per processar grans volums de dades. Són utilitzats per a la col·laboració i exploració interactiva des de notebooks de Databricks. Aquests clústers estan destinats a ser utilitzats per múltiples usuaris de manera col·laborativa. Els usuaris poden executar cel·les de codi a quaderns de forma interactiva i realitzar anàlisi exploratòria de dades.
- **Job Clúster:** És un clúster temporal utilitzat específicament per executar una feina o tasca en particular. Quan el treball es completa, el clúster es pot tancar, el que ajuda a optimitzar els recursos i controlar costos. Aquests clústers executen feines com Spark o un flux de treball d'ETL.
- **Storage Account:** Un compte d'emmagatzematge a Azure és un servei que proporciona emmagatzematge al núvol altament disponible i durador. S'utilitza per emmagatzemar dades com blobs, fitxers, taules i cues. És aquí on s'emmagatzema les metadades generades per l'unity catalog
- **Delta Lake:** És una tecnologia de gestió de dades desenvolupada per Databricks. Es basa en el format d'emmagatzematge de dades parquet i afegeix una capa de control de versions transaccional, cosa que permet l'administració de dades a nivell de taula amb transaccions atòmiques i confiabilitat en entorns de big data.
- **Azure Data Lake Storage:** És un servei d'emmagatzematge al núvol optimitzat per a anàlisi de big data. Ofereix escalabilitat massiva, seguretat millorada i capacitat per emmagatzemar i analitzar dades no estructurades i estructurades a gran escala. S'integra amb eines i serveis d'anàlisi de dades a Azure.

Als entorns de Pre i Pro els clúster estaran classificats per talles small, medium, large. Els Clústers seran efímers el que vol dir que es crearan i destruiran acabat el procés o tasca. Cal recordar que si trobem projectes que requereixin una demanda més gran de recursos (Memòria o CPU) es podrà disponibilitat de manera particular mitjançant la definició del Workflow.

Aquests talles estan disponibles a elecció per als entorns de DES, PRE i PRO.

### Cloud Públic:

Talla S – Databricks SaaS					
Clusters - Microsoft Azure i West Europe					
Identificad or d'instància	Nombre Clusters	Databricks runtime	Worker Type	Workers	Driver type
Capa Aplicacions					
1	1	13.3 LTS (Scala 2.12, Spark 3.4.1)	Memory : 14Gb Cores : 4	Min: 2 Max:8	Memory:14Gb Cores: 4

Talla M – Databricks SaaS					
Clusters - Microsoft Azure i West Europe					
Identificad or d'instància	Nombre Clusters	Databricks runtime	Worker Type	Workers	Driver type
Capa Aplicacions					
2	1	13.3 LTS (Scala 2.12, Spark 3.4.1)	Memory : 28Gb Cores : 8	Min: 2 Max:8	Memory:28Gb Cores: 8

Talla L – Databricks SaaS					
Clusters - Microsoft Azure i West Europe					
Identificad or d'instància	Nombre Clusters	Databricks runtime	Worker Type	Workes	Driver type
Capa Aplicacions					
3	1	13.3 LTS (Scala 2.12, Spark 3.4.1)	Memory : 56Gb Cores : 16	Min: 2 Max:8	Memory:56Gb Cores: 16

## 2.6.4 Denodo

El prerequisits per poder donar d'alta i treballar amb el servei IaaS de Denodo 8.0 Enterprise Plus són els següents:

- Enviar a Denodo el *tenant id* del compte d'Azure de la PTD.
- Denodo retornarà un fitxer amb la llicència per tal de que sigui activat a la plataforma.

### Entorn DES

#### Cloud Públic:

Desenvolupament - IaaS					
Clusters - Microsoft Azure i West Europe					
Identificador d'instància	Nombre VM	Programari i versió	Memòria Ram i Recursos addicionals	Disc Persistent	Administrat per CPD (Si/No)
Capa Aplicacions					
1	2	Denodo 8.0 Enterprise Plus	4 vCPUs – 32 GB	256 GB	No

### Entorn Preproducció

#### Cloud Públic:

Preproducció - IaaS					
Clusters - Microsoft Azure i West Europe					
Identificador d'instància	Nombre VM	Programari i versió	Memòria Ram i Recursos addicionals	Disc Persistent	Administrat per CPD (Si/No)
Capa Aplicacions					
1	2	Denodo 8.0 Enterprise Plus	4 vCPUs – 32 GB	256 GB	No

Preproducció - SaaS					
Base de Dades - Microsoft Azure i West Europe					
Identificador d'instància	Nombre BDD	Programari i versió	Memòria Ram i Recursos addicionals	Disc Persistent	Administrat per CPD (Si/No)
Capa Aplicacions					
2	1	Azure SQL	Min: 1 vCPUs – Max: 2 vCPUs RAM: 6 GiB	128 GB	No

## Entorn Producció

### Cloud Públic:

Producció - IaaS					
Clusters - Microsoft Azure i West Europe					
Identificador d'instància	Nombre VM	Programari i versió	Memòria Ram i Recursos addicionals	Disc Persistent	Administrat per CPD (Si/No)
Capa Aplicacions					
3 2	2	Denodo 8.0 Enterprise Plus	4 vCPUs – 32 GB	256 GB	No

Producció - SaaS					
Base de Dades - Microsoft Azure i West Europe					
Identificador d'instància	Nombre BDD	Programari i versió	Memòria Ram i Recursos addicionals	Disc Persistent	Administrat per CPD (Si/No)
Capa Aplicacions					
4	1	Azure SQL	Min: 1 vCPUs – Max: 2 vCPUs RAM: 6 GiB	128 GB	No

### 2.6.5 Altres recursos a desplegar

Per cobrir altres necessitats serà necessari el desplegament del següents recursos:

Recurs	Descripció	Localització
Azure Key Vault	Gestiona i mantén claus de l'aplicació, entre elles les de xifratge de les dades que s'emmagatzemen a Azure i MongoDB Atlas.	Azure.
Azure Active Directory	Emprar el que ja té CTTI a Azure.	Azure
VNet Private Links	Links privats entre les VNets d' Azure i les de MongoDB Atlas.	Azure
Azure Kubernetes Services	Entorn totalment gestionat que permet executar microserveis i aplicacions en contenidor en una plataforma sense servidor.	Azure



## 2.6.6 Altres dades rellevants pel desplegament

Dada a proporcionar	Opcions / Detall
Xarxes d'accés	<input checked="" type="checkbox"/> Internet <input checked="" type="checkbox"/> Intranet <input checked="" type="checkbox"/> Extranet
Servei transversal SMTP	<input checked="" type="checkbox"/> Si <input type="checkbox"/> No
ProxyPass (Sortida a Internet)	<input checked="" type="checkbox"/> Si <input type="checkbox"/> No
Altres serveis tècnics utilitzats	-

### Justificacions de les decisions de la vista de desplegament:

Decisió	Justificació
Entorn DES, PRE i PRO	Entorn de desenvolupament, preproductiu i productiu.
Desplegament al cloud públic	Inicialment, es considera desplegar al Cloud públic de Microsoft Azure.

## 2.6.7 Vista Operacional

Gestió de logs	Opcions / Detall
Indicar l'activitat a registrar	S'enregistrarà com a mínim: <ul style="list-style-type: none"> <li>• Intents d'accés al sistema (exitosos i fallits).</li> <li>• Altes, baixes i canvis en els permisos dels usuaris.</li> <li>• Canvis en la configuració de l'aplicació.</li> <li>• Accessos i modificacions a dades sensibles.</li> </ul>
Política de rotació i retenció dels Logs.	El període de retenció de logs serà de 6 mesos disponibles en tot moment i de 12 mesos disponibles sota demanda, el que implica un total de 18 mesos de retenció.
Ubicació dels logs	Els propis de la plataforma.

Sondes	Descripció
<b>Detall de les Sondes</b>	Azure, MongoDB Atlas i Databricks implementen eines que permeten monitoritzar la plataforma de dades. S'han d'integrar amb Talaia.

Polítiques de retenció	Descripció
Identificar quina de les polítiques s'ajusta més al que es requereix.	<input type="checkbox"/> Estàndard <input type="checkbox"/> Avançada <input type="checkbox"/> Especial <u><i>Més informació respecte a cada Política</i></u>

### 3. Perspectives Transversals

#### 3.1 Seguretat

Dada a proporcionar	Opcions / Detall
<b>Mesures de seguretat bàsiques de l'Agència de Ciberseguretat</b>	<b><u>Mesures de seguretat.</u></b> <input checked="" type="checkbox"/> S'ha llegit i es tindran en compte les mesures de seguretat vigents a l'hora d'implementar l'arquitectura del servei / solució.
<b>Sistema d'autenticació</b>	<input checked="" type="checkbox"/> Usuari Intern (Gicar) <input type="checkbox"/> Usuari Extern (VÀLid) <input type="checkbox"/> Accés Híbrid (Gicar i VÀLid) <input type="checkbox"/> No requereix Autenticació <input checked="" type="checkbox"/> Altres: usuari/password (usuaris tècnics de BDD de MongoDB Atlas i Denodo i usuaris dels servidors SFTP), Microsoft Entra ID del CTTI per als usuaris d'Azure Databricks, aplicacions empresarials per al usuaris màquina de Databricks.  Els usuaris SFTP seran engabiats i empraran usuari/password per autenticar-se.  Els passwords o secrets dels usuaris tècnics seran emmagatzemats a Azure Key Vault.
<b>Modalitat d'integració amb Gicar</b>	<input type="checkbox"/> SiteMinder <input type="checkbox"/> Agent de Shibboleth <input checked="" type="checkbox"/> SAML Out of the box <input type="checkbox"/> Canigó SAML2 <input type="checkbox"/> ADFS-GICAR <input type="checkbox"/> AD, LDAP, o BBDD aprovisionada per GICAR <input checked="" type="checkbox"/> Altres:
<b>Usuaris especials</b>	Un usuari amb permisos de Global Admin de Microsoft Entra ID per poder activar el Databricks Unity Catalog de la PTD. Posteriorment l'usuari ha de crear el administradors de comptes de Databricks.
<b>Xifrat infraestructura</b>	Les dades en repòs s'han de xifrar als magatzems d'Azure definits per a Databricks, Denodo i MongoDB Atlas, fent servir Azure Key Vault com a repositori de claus de xifratge, que seran gestionades pel CTTI.  <a href="https://docs.microsoft.com/es-es/azure/security/fundamentals/encryption-overview">https://docs.microsoft.com/es-es/azure/security/fundamentals/encryption-overview</a>  Les dades en transit sempre aniran xifrades amb protocols segurs, TLS i SFTP.

## Certificacions



La categoria ENS és ALTA. La molta criticitat de les dades implica que aquest serà ALT a nivell ENS i per tant la categoria de la plataforma es considera ALTA segons article 40 i Annex I de l'ENS.

**Microsoft Azure** està certificat en el compliment de l'ENS, nivell ALT, com s'indica a la pàgina Web de l'ENS (<https://gobernanza.ccn-cert.cni.es/certificados>). Els Sistemes d'Informació que suporten els serveis de Microsoft Azure són enumerats a la pàgina de l'Annex del Certificat, que es pot veure al següent document:



ES\_Certificado de conformidad ENS M

**MongoDB Atlas** està certificat, entre d'altres, en: ISO 27001/2013, CSA Start Level 2 i SOC 2 Type 2, però no en l'ENS, encara que en un correu enviat a Minsait indiquen que estan realitzant els tràmits adients.



Certificat ISO 27001, 2013



MongoDB - 2022 CSA STAR Certificati



MongoDB Atlas - SOC 2 + HITRUST M

En el següent arxiu es pot veure una comparativa realitzada per Minsait entre SOC 2 Type 2 i l'ENS:



mapping-soc-2-ENS  
.xlsx

**Denodo** està certificat en ISO 27001/2013, però no s'ha trobat el certificat corresponent. No està certificat en l'ENS, però indiquen que estan en tràmits, com es pot veure al correu adjunt:



cumplimiento  
Denodo certificado

**Informatica** està certificat en el ENS:



Informatica\_Certificado de conformidad EN

### 3.2 Rendiment i escalabilitat

Dada a proporcionar	Opcions / Detall
Requeriments de rendiment continuat i davant pics	
Mesures adoptades per tal d'assolir el rendiment necessari	L'aplicació està preparada per l'escalabilitat horitzontal? <input checked="" type="checkbox"/> Si <input type="checkbox"/> No  Perquè: El sistema està desplegat al núvol de Azure i tindrà opció a comptar amb recursos auto escalables.

### 3.3 Disponibilitat

Dada a proporcionar	Opcions / Detall
RTO del Sistema	Temps que pot estar el negoci amb el servei aturat. <input type="checkbox"/> 2 hores <input checked="" type="checkbox"/> Entre 2 i 24 hores <input type="checkbox"/> Més de 24 hores <input type="checkbox"/> Altres:
RPO Punt de recuperació Objectiu	En cas d'incidència quin es desitja que sigui el punt de recuperació:  <input type="checkbox"/> <b>Zero:</b> No hi ha pèrdua d'informació, el sistema de recolzament ha de tenir exactament la mateixa que hi havia abans de l'incident. <input checked="" type="checkbox"/> <b>Darrer Backup:</b> En cas d'incident, el sistema es recupera amb l'últim backup conegut. <input type="checkbox"/> <b>Altres:</b>
Definir horari de servei habitual	<input type="checkbox"/> Laboral (12x5) <input checked="" type="checkbox"/> Continu (24x7) <input type="checkbox"/> Altres

Afectació per la indisponibilitat d'entitats externes:

Sistema extern	Descripció

### 3.4 Localització

No aplica

<b>Dada proporcionar</b>	<b>Opcions / Detall</b>
<b>Idiomes que suporta el sistema</b>	<input checked="" type="checkbox"/> Català <input checked="" type="checkbox"/> Aranès <input checked="" type="checkbox"/> Castellà <input type="checkbox"/> Anglès <input type="checkbox"/> Francès <input type="checkbox"/> Altres:
<b>Definir com es resol multilingüe</b>	<input type="checkbox"/> Mitjançant la rèplica d'un conjunt de pàgines per a cada idioma. <input checked="" type="checkbox"/> Mitjançant un únic conjunt de pàgines que obtenen diferents literals de forma externa segons l'idioma. <input checked="" type="checkbox"/> Altres: Col·lecció dades de referència.

## 4. Informació específica pel projecte d'aprovisionament

### 4.1 Informació relativa al context

No aplica

### 4.2 Informació relativa al SIC

Dada a proporcionar	Opcions / Detall
Entorns a gestionar pel SIC	Desenvolupament, Preproducció i Producció.
Organització de branques	dev, pre, màster
Artefactes	Es generaran els artefactes corresponents els projectes desenvolupats.

### 4.3. Informació relativa a xarxes i dominis DNS

A continuació es llista la informació relativa als entorns de preproducció i producció. La configuració és pràcticament idèntica, per tant s'agrupa en la mateixa taula i només es distingirà quan sigui menester.

Entorn	Origen	Host Destí	IP Destí	Port	Protocol	Integració amb
PRE/PRO/DES	Xarxa CTTI	PTD via Azure ExpressRoute	DES: X.X.X.X/XX  PRE: 172.29.22.0/24  PRO: 172.29.23.0/24		TCP	Databricks Denodo Azure Container Instances MongoDB Atlas  A MongoDB Atlas s'hi accedirà via una VNET a través del Private Link configurat.

Entorn	Origen	Host Destí	IP Destí	Port	Protocol	Integració amb
PRE/PRO/DES	Azure PTD	<b>MongoDB Cluster</b> via Private Link <b>Atlas</b> Azure	<b>DES:</b> Azure <input type="checkbox"/> X.X.X.X/XX MongoDB <input type="checkbox"/> Intern  <b>PRE:</b> Azure <input type="checkbox"/> 172.29.22.0/24 MongoDB <input type="checkbox"/> Intern  <b>PRO:</b> Azure <input type="checkbox"/> 172.29.23.0/24 MongoDB <input type="checkbox"/> Intern	1024 a 1173	TCP	Clúster de MongoDB Atlas.
	Informatica AXON/EDC/IDQ	<b>PTD</b> via Private Link Azure	<b>PRE:</b> Informatica <input type="checkbox"/> X.X.X.X/XX PTD <input type="checkbox"/> 172.29.22.0/24  <b>PRO:</b> Informatica <input type="checkbox"/> X.X.X.X/XX PTD <input type="checkbox"/> 172.29.23.0/24		TCP	



	Mongo DB Atlas Cluster	Azure Key Vault	Intern Azure	TLS	Adreces de MongoDB Atlas des de les quals es poden fer peticions a Azure Key Vault:  3.92.113.229 3.208.110.31 3.211.96.35 3.212.79.116 3.214.203.147 3.215.10.168 3.215.143.88 3.232.182.22 18.214.178.145 18.235.30.157 18.235.48.235 18.235.145.62 34.193.91.42 34.193.242.51 34.194.7.70 34.196.80.204 34.196.151.229 34.200.66.236 34.235.52.68 34.236.228.98 34.237.40.31 34.238.35.12 35.153.40.82 35.169.184.216 35.171.106.60 35.173.54.44 35.174.179.65 35.174.230.146 35.175.93.3 35.175.94.38 35.175.95.59 44.206.200.18 44.207.9.197 44.207.12.57 50.19.91.100 52.7.232.43 52.71.233.234 52.73.214.87 52.87.98.128 52.203.106.167 54.145.247.111 54.163.55.77 54.167.217.16 100.26.2.217 107.20.0.247
--	------------------------	-----------------	--------------	-----	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Entorn	Origen	Host Destí	IP Destí	Port	Protocol	Integració amb
						107.20.107.166 107.22.44.69
	PTD PRE	Eventhub PRE (preproduccio.event hub.intranet.gencat. cat)	10.53.194.11	6000 6001 6002 6003	TLS	Kafka Broker Cluster 6000 -> LEVHIFX60:6001, LEVHIFX61:6002, LEVHIFX62 :6003 6001 -> LEVHIFX60:6001 6002 -> LEVHIFX61:6002 6003 -> LEVHIFX62:6003  La connexió entre ambdós <i>endpoints</i> es realitzarà a través d'Azure ExpressRoute.  L'autenticació es realitzarà mitjançant mTLS.

Entorn	Origen	Host Destí	IP Destí	Port	Protocol	Integració amb
	PTD PRO	Eventhub PRO (eventhub.intranet.gencat.cat)	10.52.194.10	6000 6001 6002 6003	TLS	<p>Kafka Broker Cluster 6000 -&gt; LEVHIFT60:6001, LEVHIFT61:6002, LEVHIFT62 :6003 6001 -&gt; LEVHIFT60:6001 6002 -&gt; LEVHIFT61:6002 6003 -&gt; LEVHIFT62:6003</p> <p>La connexió entre ambdós <i>endpoints</i> es realitzarà a través d'Azure ExpressRoute.</p> <p>L'autenticació es realitzarà mitjançant mTLS.</p>
	PTD	EDC/AXON	172.31.246.32/27	443	TLS	<p>S'ha de crear un VNet Peering entre la VNet d'EDC/AXON i la de la Plataforma de Dades.</p> <p>Per a PRE, la configuració de la VNet d'EDC/AXON és la següent (manca la de PRO, que encara no existeix):</p> <p><b>Subscripció:</b> Gencat CTTI preproduccio 8e572353-4a3f-40cf-8698-cea015266df3 <b>RG:</b> tdd-common-rg <b>Vnet:</b> tdd-pre-vnet <b>Address Space:</b> 172.31.246.32/27</p>

Entorn	Origen	Host Destí	IP Destí	Port	Protocol	Integració amb
INT	PowerBI (10.51.111.209)	PTD	172.29.22.0/24	443	JDBC	Databricks
PRE	PowerBI (10.51.193.41)	PTD	172.29.22.0/24	443	JDBC	Databricks
PRO	PowerBI (10.50.193.38)	PTD	172.29.23.0/24	443	JDBC	Databricks

Entorn	Origen	Host Destí	IP Destí	Port	Protocol	Integració amb
PRE	Databricks PTD (172.29.22.64/27)	CPD2/REU	10.51.228.20	1591	TCP	Origen de dades de REU PRE.
PRO	Databricks PTD (172.29.23.64/27)	CPD2/REU	10.50.228.24	1591	TCP	Origen de dades de REU PRO.

Font de les regles per MongoDB Atlas: [\*Security Features and Setup — MongoDB Atlas\*](#)

S'han de configurar les següents necessitats de connectivitat:

- Crear d'un enllaç tipus Private Link entre la VNET d'Azure a on es troba desplegat el Databricks i la VNET de MongoDB Atlas.

Dada proporcionar	a	Opcions / Detall
Dominis DNS		PRE: preproduccio.ptd.ctti.intranet.gencat.cat 172.29.22.85 preproduccio.api.ptd.ctti.intranet.gencat.cat 172.29.22.85 preproduccio.govern.ptd.ctti.intranet.gencat.cat 172.29.22.85 preproduccio.denodo.ptd.ctti.intranet.gencat.cat

	<p>           172.29.22.134 adb-7139081274934771.11.azuredatabricks.net            172.29.22.134 adb-dp-7139081274934771.11.azuredatabricks.net            172.29.22.133 adb-5780738585396366.6.azuredatabricks.net            172.29.22.133 adb-dp-5780738585396366.6.azuredatabricks.net            172.29.22.165 adb-3807468366949474.14.azuredatabricks.net            172.29.22.165 adb-dp-3807468366949474.14.azuredatabricks.net            172.29.22.167 adb-589894607047291.11.azuredatabricks.net            172.29.22.167 adb-dp-589894607047291.11.azuredatabricks.net            172.29.22.168 adb-2316150342427455.15.azuredatabricks.net            172.29.22.168 adb-dp-2316150342427455.15.azuredatabricks.net            —            PRO:            ptd.ctti.intranet.gencat.cat            172.29.23.85 api.ptd.ctti.intranet.gencat.cat            172.29.23.85 govern.ptd.ctti.intranet.gencat.cat            172.29.23.85 denodo.ptd.ctti.intranet.gencat.cat            —            172.29.23.133 adb-3160202689475972.12.azuredatabricks.net            172.29.23.133 adb-dp-3160202689475972.12.azuredatabricks.net            172.29.23.134 adb-1917381938162982.2.azuredatabricks.net            172.29.23.134 adb-dp-1917381938162982.2.azuredatabricks.net            172.29.23.165 adb-415924557256886.6.azuredatabricks.net            172.29.23.165 adb-dp-415924557256886.6.azuredatabricks.net            172.29.23.167 adb-4258240250330622.2.azuredatabricks.net            172.29.23.167 adb-dp-4258240250330622.2.azuredatabricks.net            172.29.23.168 adb-52672067831854.14.azuredatabricks.net            172.29.23.168 adb-dp-52672067831854.14.azuredatabricks.net         </p>
<b>Urls a assegurar amb Gicar</b>	<p>           MongoDB Atlas: <a href="https://cloud.mongodb.com">https://cloud.mongodb.com</a>            —            PRE:  <a href="https://preproduccio.govern.ptd.ctti.intranet.gencat.cat/*">https://preproduccio.govern.ptd.ctti.intranet.gencat.cat/*</a>  <a href="https://preproduccio.denodo.ptd.ctti.intranet.gencat.cat/*">https://preproduccio.denodo.ptd.ctti.intranet.gencat.cat/*</a>            —            PRO:  <a href="https://govern.ptd.ctti.intranet.gencat.cat/*">https://govern.ptd.ctti.intranet.gencat.cat/*</a>  <a href="https://denodo.ptd.ctti.intranet.gencat.cat/*">https://denodo.ptd.ctti.intranet.gencat.cat/*</a> </p>
<b>CIDR</b>	<p> <b>DES:</b> X.X.X.X/XX (direccionament CTTI) i 10.10.0.0/16 (direccionament intern plataforma)            —  <b>PRE:</b> 172.29.22.0/24 (direccionament CTTI) i 10.10.0.0/16 (direccionament intern plataforma)            —  <b>PRO:</b> 172.29.23.0/24 (direccionament CTTI) i 10.10.0.0/16 (direccionament intern plataforma)         </p>
<b>Azure ExpressRoute</b>	<p> <b>DES:</b> X.X.X.X/XX  <b>PRE:</b> 172.29.22.0/24  <b>PRO:</b> 172.29.23.0/24         </p>

## 4.4 Informació relativa a l'aprovisionament d'Infraestructura

[No aplica]

Capa	Identificador d'instància	Explicació

## 5. Referències

[1]: MongoDB Atlas. Microsoft Azure. URL:

<https://docs.atlas.mongodb.com/reference/microsoft-azure/>

[2]: Databricks. URL:

<https://learn.microsoft.com/en-us/azure/databricks/administration-guide/#establish-first-account-admin>

[3]: Certificacions Informàtica. URL: <https://www.informatica.com/trust-center/certifications-assessments-standards.html#fbid=uEbHlz0jUn4>