Mètodes d'Ingesta de Dades a la PTD

Plataforma Transversal de Dades

Febrer 2026

Finançat per la Unió Europea NextGenerationEU
GOBIERNO DE ESPAÑA
Plan de Recuperación, Transformación y Resiliencia
Next Generation Catalunya
Generalitat de Catalunya

1

Tasques preliminars a escollir un mètode d’Ingesta

Tasques preliminars a escollir un mètode d'Ingesta

- Identificar el **propietari** de les dades.
- Identificar les **parts interessades** que participaran al projecte:
  - Departament i proveïdors implicats.
  - Referents de dades tècnics.
  - Referents funcionals.
  - Referents de govern de les dades (llistat disponible 'Conèixer el govern de les dades')
- Identificar **domini/subdomini** de les dades.
  - Agrupació temàtica nivell domini | Dades obertes de Catalunya
  - Agrupació temàtica nivell subdomini | Dades obertes de Catalunya
- Identificar el **centre de cost** del promotor del cas d'ús i nom del projecte (FinOps).
- Determinar el grau de **qualitat de les dades** d'origen i les accions per a millorar-la.
- Assegurar l'acompliment de la regulació relativa a **dades sensibles.**

2

Fases pel disseny d’un procés d’ingesta PTD

Fases pel disseny d’un procés d’ingesta PTD

La implantació d’una solució d’ingesta és un procés complex, però això és recomanable tenir **un pla predefinit** (fulla de ruta).

1

Anàlisis previ

- Requisits de Qualitat
- Anàlisi dades i metadades
- Tipus d’Extracció
- Triggers i periodicitat
- Estructura (entitats)

2

Definició de la solució

- Tipus d’Ingesta
- Regles QA
- Mecanisme de Publicació
- Nomenclatura (Govern)
- Tractament dades sensibles (PII)

3

Detalls tècnics

- Consum i disponibilització
- Report i notificacions
- Consum Cluster (finOps)
- Emmagatzemament
- Accés i permisos

Fase 1 - Anàlisis Previ

En aquesta fase es defineixen els criteris i requisits que hauran de complir les dades que s'ingesten.

Requisits de Qualitat	Nivell de qualitat actual –vs- nivell desitjat Indicadors de qualitat. Nivell de qualitat de les fonts.
Anàlisi dades i metadades	Tipus i formats. Mostres de dades. Identificar problemes o inconsistències.
Fonts de dades	Tècniques i eines d'extracció. Volumetria. Fluxos de dades. Tipus de font (base de dades, fitxer, API...)
***Trigers* i periodicitat**	Cada quant s'ha de realitzar l'extracció. Quins esdeveniments (triggers) desencadenen.
Estructura (entitats)	Entitats. Regles de negoci. Relacions entre entitats. Estructura de les taules destí.

Fase 2 - Definició de la solució

Aquesta fase assegura que les dades arribin correctament i que siguin gestionades sota criteris clars.

Tipus d'Ingesta	Es defineix com s'incorporaran les dades dins de la PTD
Regles QA	Regles de validació, remediació i enriquiment de les dades.
Mecanisme de Publicació	Com i on es publiquen les dades un cop processades per a ser consumides.
Nomenclatura (Govern)	Nom de taules, nom de camps i entitats. Normes de gestió del cicle de vida de les dades.
Tractament de dades sensibles (PII)	S'identifiquen les dades sensibles i s'apliquen mesures (encriptació, anonimització, control d'accés…)

Fase 3 - Detall tècnic

L'objectiu d'aquesta fase és garantir un accés eficient, segur i controlat a les dades.

Consum i disponibilització	Com es posen les dades a disposició dels usuaris (API, dades obertes, exportació...)
Report i notificacions	Qui ha de ser notificat i com. Alertes automàtiques (email)
Consum Cluster (finOps)	Ús i cost del clúster per projecte.
Emmagatzemament	Tipus d'emmagatzematge (temporal –vs- històric). Configuració sistemes d'emmagatzematge
Accés i permisos	Qui pot accedir. Rols i polítiques. Autenticació.

3

Eines d’ingesta disponibles

Eines d’ingesta disponibles:

Eines pròpies de Databricks

- **Lakeflow** - Connect
- **Lakeflow** - Spark Declarative Pipelines (SDP)
- **Lakeflow** - Jobs
- **Lakehouse** Federation

Eines disponibles a la PTD integrades amb Databricks

- Fivetran
- Denodo
- Eventhub
- ITQ
- Plantilles

4

Eines pròpies de Databricks

4.1

Eines pròpies de Databricks

Lakeflow Connect

Eines pròpies de Databricks

Lakeflow Connect

Lakeflow Connect és una funcionalitat de Databricks que permet ingestar dades a bronze de manera senzilla i automatitzada des de múltiples fonts cap a Delta Lake, **sense necessitat d'escriure codi** complex. Està pensada per **simplificar la càrrega inicial de dades** i garantir que aquestes arribin al Lakehouse amb consistència i governança.

Permet extreure dades de diverses fonts, transformar-les i carregar-les en un Delta Lake.

- "Jobs & Pipelines" > "Ingestion Pipeline"

Llegeix dades des de:

- Fitxers locals i sistemes d'arxius distribuïts.
- Bases de dades relacionals.
- APIs i Web Services
- Connectors nadius SaaS (Salesforce, SQL Server, Google Analytics Raw Data, SAP Business Data Cloud, Workday Reports, ServiceNow)
- Connectors de partners (OneDrive, Google Drive, Jira, GitHub, etc)

Lakeflow Connect:
https://docs.databricks.com/aws/en/ingestion/lakeflow-connect/

CONTRES	PROS
Els nous connectors SaaS 1er passen pel mode Beta, com el de SharePoint, el que pot afectar a l'estabilitat de la integració.	No-code / Low-code: Ideal per equips no tècnics doncs permet configurar ingestes sense escriure codi.
Menor flexibilitat que un Notebook.	Compatible amb fitxers locals, bases de dades relacionals, APIs, connectors SaaS i serveis de Partners.
No és òptim per transformacions avançades; només per ingestes simples.	Integració nativa amb Delta Lake: Inclou funcionalitats com control de versions, rollback i time travel.
Si no existeix connector nadiu, cal desenvolupar solucions alternatives.	Compatible amb Unity Catalog per seguretat, lineage i control d'accés.

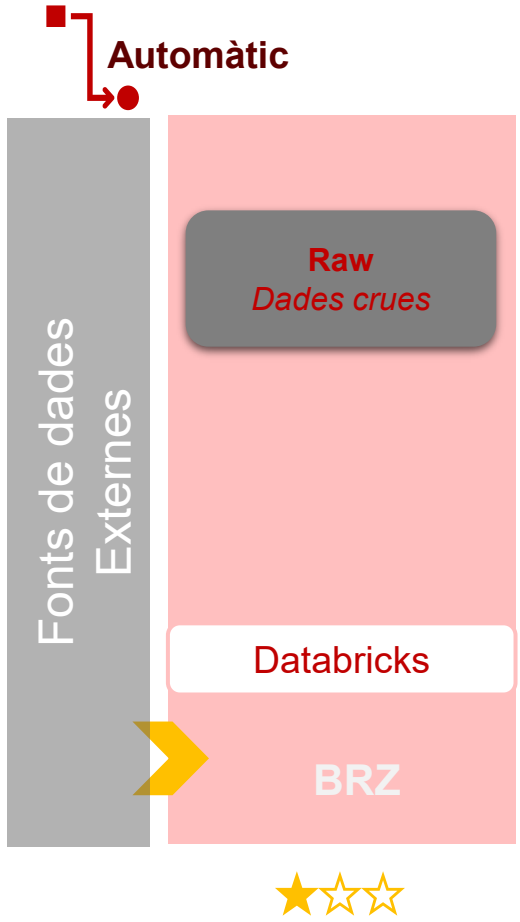
Eines pròpies de Databricks

Lakeflow Connect

Consell:

Fes servir Ingestion Pipeline de Lakeflow Connect quan necessitis una **ingesta inicial o recurrent** de dades des de múltiples fonts cap al teu Delta Lake de manera **ràpida, automatitzada i** amb **mínima complexitat** de codi.

- Projectes amb alta diversitat de fonts (fitxers locals, bases de dades, APIs, connectors SaaS).
- Fase d'onboarding de dades per carregar grans volums a la capa bronze sense invertir temps en ETL manual per disponibilitzar-les a la plataforma.
- Necessitat de governança i consistència des del primer moment, assegurant traçabilitat i qualitat.
- Equip amb menys experiència en desenvolupament Spark o quan vols reduir dependència de scripts personalitzats.



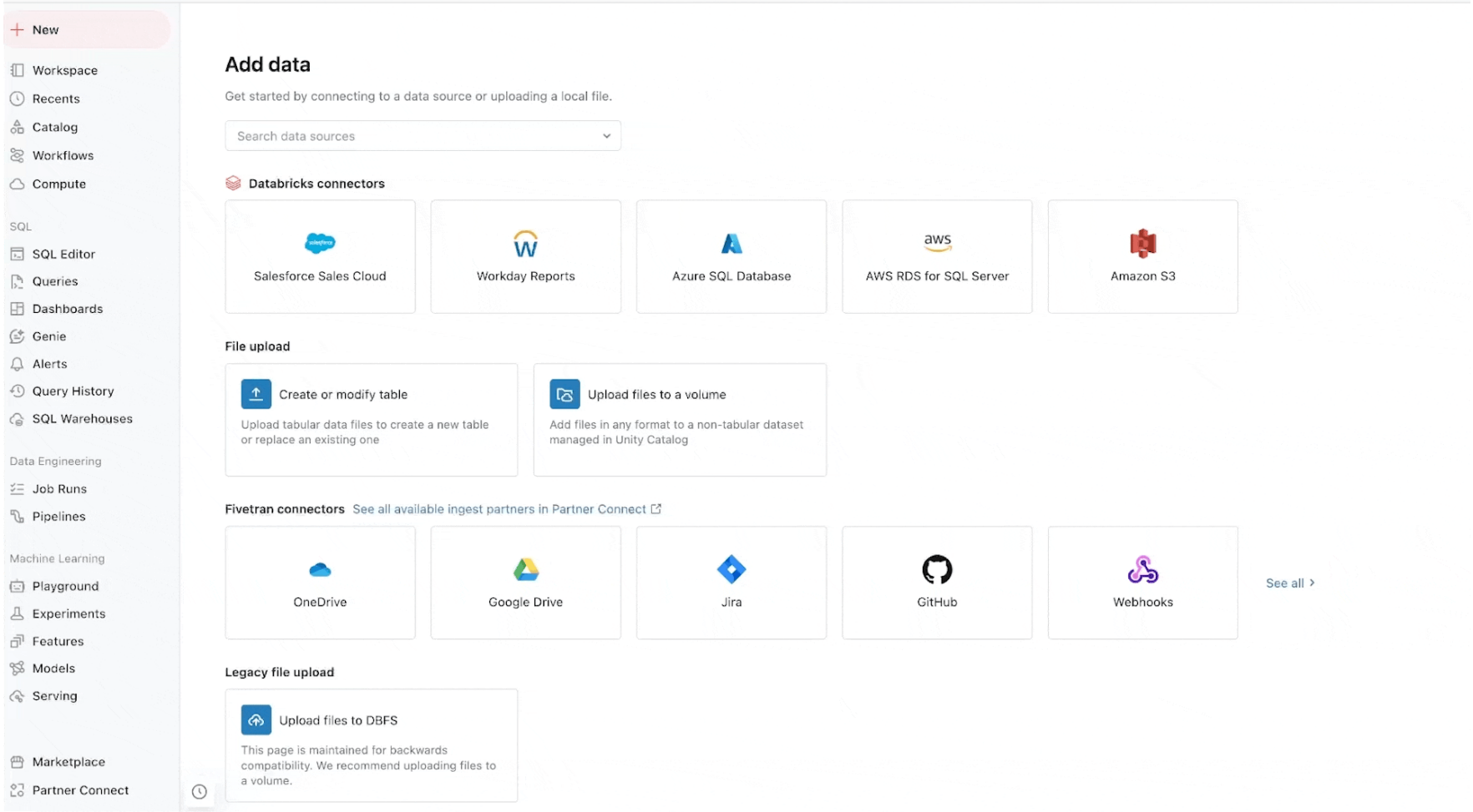
Eines pròpies de Databricks

Lakeflow Connect

Lakeflow Designer s'utilitza per construir pipelines ETL de manera visual i sense codi.

- No-code / Low-code
- Integración nativa con Unity Catalog

https://www.databricks.com/resources/demos/tours/data-engineering/introducing-lakeflow-designer



4.2

Eines pròpies de Databricks

Lakeflow Spark Declarative Pipelines (SDP)

Eines pròpies de Databricks

Lakeflow Spark Declarative Pipelines (SDP)

Aquest tipus és adequat per a processos **ETL declaratius i ingestes contínues**.

SDP és la funcionalitat actual de Databricks que substitueix Delta Live Tables (DLT). SDP permet crear i **gestionar pipelines de dades de manera declarativa**, aprofitant la potència de Spark i la integració amb Delta Lake, amb governança i seguretat integrada.

És ideal per **processos ETL robustos**, transformacions complexes i ingestes en temps real o per lots.

- "Jobs & Pipelines" > "ETL Pipeline"

Llegeix dades des de:

- Fitxers (CSV, JSON, Parquet)
- Kafka
- BDD JDBC (MySQL, SQL Server, PostgreSQL)
- Autoloader

Lakeflow Spark Declarative Pipelines (SDP):
https://learn.microsoft.com/es-es/azure/databricks/ldp/

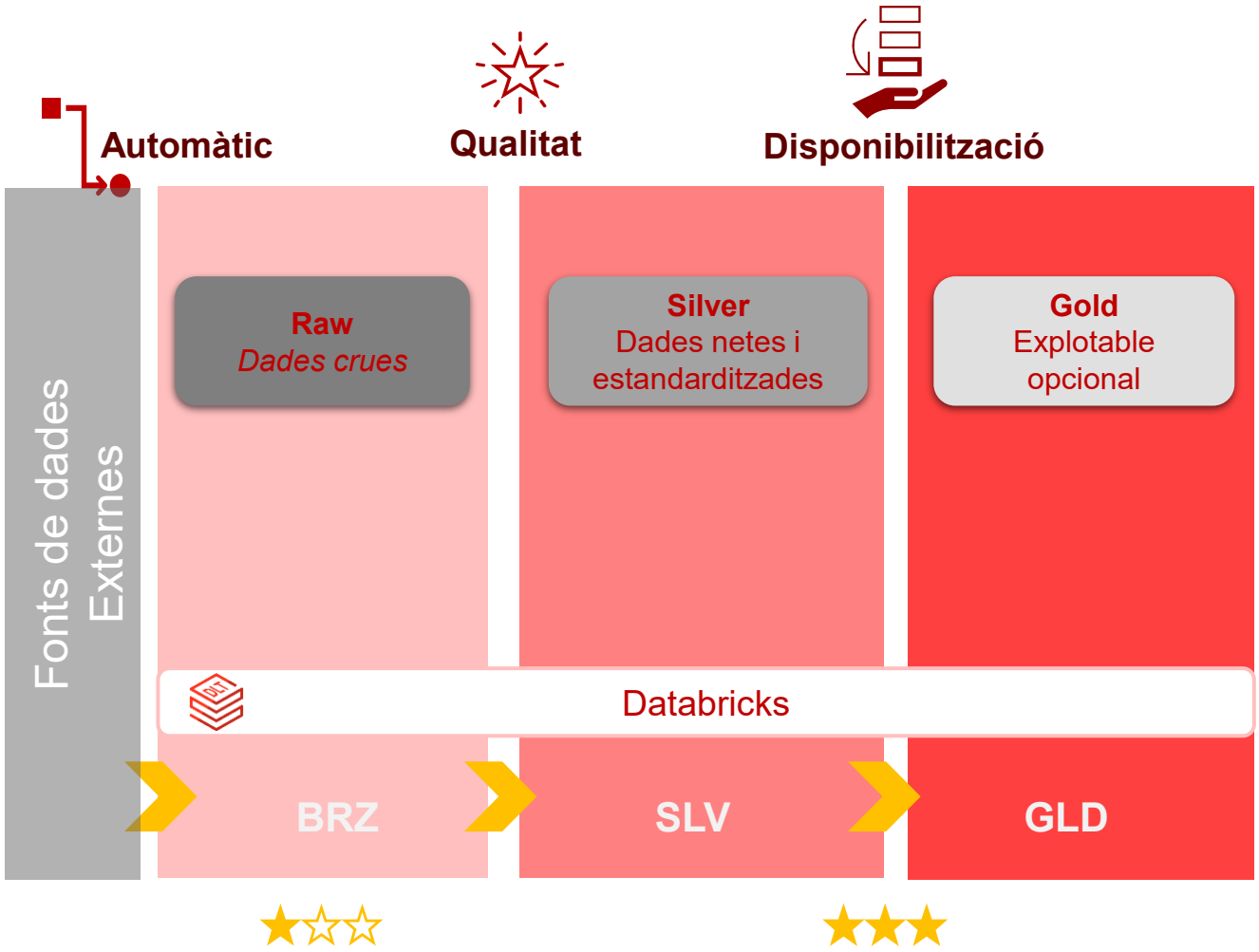
CONTRES	PROS
Si no es controlen els recursos, es pot incrementar la **despesa en Streaming**.	**Pipelines declaratius** i fiables per ETL i ingestes
Menys flexibilitat per a lògica complexa que els Jobs Ad-hoc.	Integració amb **Unity Catalog per una Governança** centralitzada i segura.
Requereix **versions recents del runtime** per funcionalitats avançades.	Ús d'**Expectations per definir regles de qualitat**.
Pot requerir **coneixement de SQL/Python** per transformacions avançades.	**Compatible amb Unity Catalog** per seguretat i lineage.

Eines pròpies de Databricks

Lakeflow Spark Declarative Pipelines (SDP)

Consell:
Utilitza **Lakeflow Spark Declarative Pipelines (SDP)** quan **necessitis pipelines fiables i automatitzats** per ingestes contínues o en temps real amb capacitat d'aplicar **transformacions declaratives** i validar la qualitat de les dades mitjançant **Expectations**.

Evita SDP si només necessites una **ingesta molt senzilla i puntual** (en aquest cas és millor Ingestion Pipeline), o si requereixes **lògiques molt personalitzades** que no encaixen en un model declaratiu.



4.3

Eines pròpies de Databricks
Lakeflow Job

Eines pròpies de Databricks

Lakeflow Job

Lakeflow Jobs és el **motor d'orquestració** de Databricks que permet **automatitzar i gestionar fluxos** de treball complexos dins del Lakehouse, definint workflows amb tasques com **ingestes, transformacions, execució de notebooks, queries SQL, pipelines declaratius (SDP) i scripts externs**. Essencial per convertir aquests pipelines i notebooks en **processos productius**, fiables i governats, evitant l'orquestració manual i reduint errors operatius.

- "Jobs & Pipelines" > "ETL Pipeline"

Serveix per:

- **Programar processos recurrents** (diaris, setmanals, per lots o en temps real).
- **Coordinar dependències** entre diferents tasques i assegurar l'ordre correcte d'execució.
- **Integrar ingestes i ETL declaratius** amb components com Machine Learning, BI o exportacions.
- **Monitorar i alertar** sobre l'estat de les execucions, amb logs detallats i notificacions.
- **Escalar workflows** en entorns distribuïts amb compute dedicat o serverless.

Lakeflow Job:
https://docs.databricks.com/aws/en/jobs/

CONTRES	PROS
Ha de muntar-ho tot. Pot requerir **coneixement avançat** per workflows molt complexos.	**Automatització completa** de workflows complexos
Risc de **cost computacional elevat** si no es dimensiona correctament.	Integració amb **pipelines SDP i ingestes Lakeflow**.
Les **UDFs de DR** no es poden executar directament en un clúster SQL	Pot aplicar **qualitat invocant les UDFs d'Spark o del catàleg comú** de la PTD.
Menys adequat per ingestes simples (millor Ingestion Pipeline de Lakeflow connect).	**Governança i seguretat integrada** amb Unity Catalog.
Depèn de **versions recents per funcionalitats avançades**.	Suport per **dependències i condicions entre tasques**.

Eines pròpies de Databricks

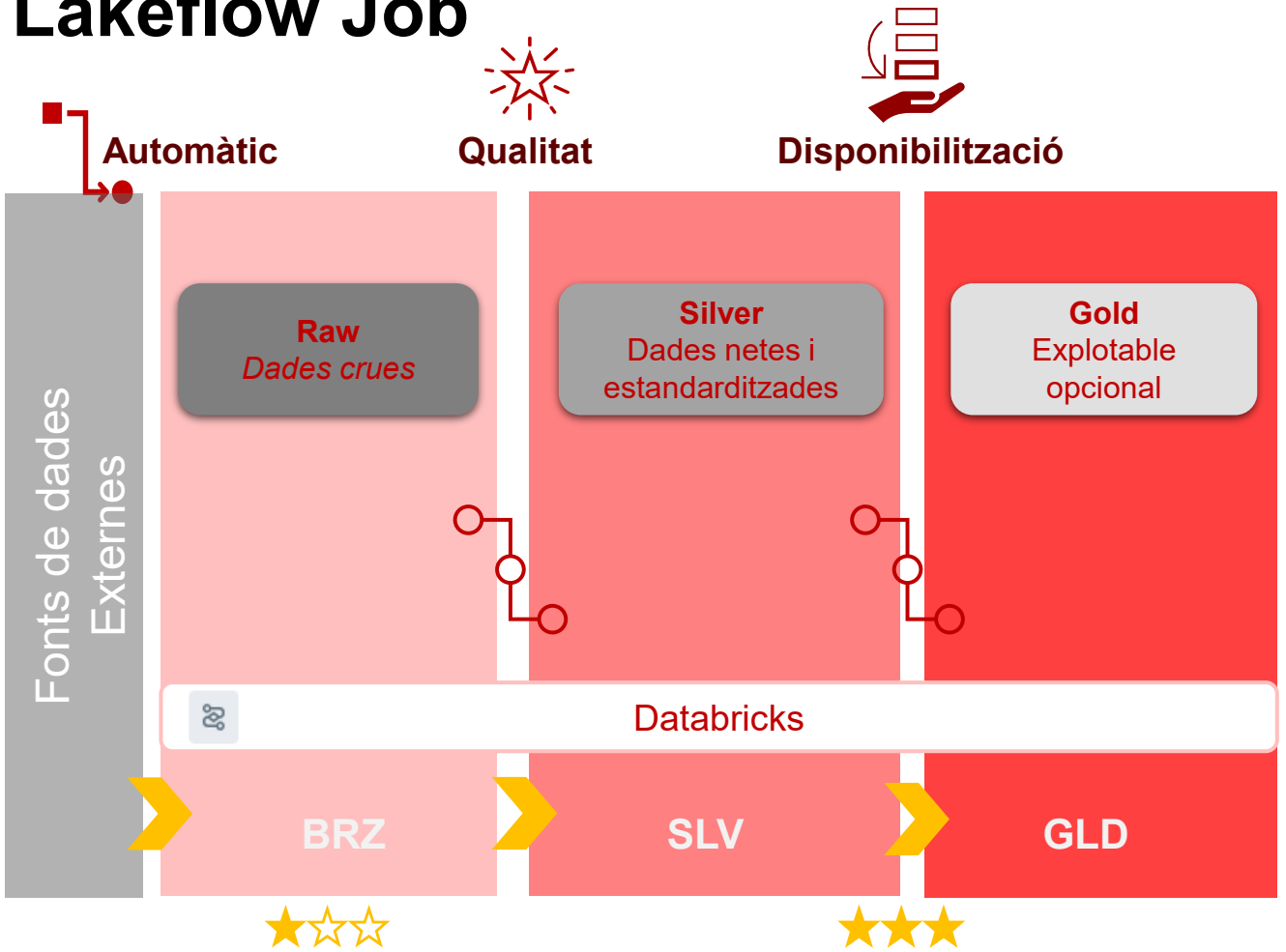
Lakeflow Job

Consell:

Fes servir Lakeflow Jobs quan necessitis orquestrar i automatitzar fluxos de treball complexos que combinin diverses tasques dins del Lakehouse. Especialment útil quan:

- Quan has de **programar processos recurrents** (ingestes, ETL, entrenament de models, exportacions).
- Quan vols **coordinar dependències** entre notebooks, pipelines declaratius (SDP), queries SQL i scripts externs.
- Quan necessites **monitoratge, alertes i governança integrada** amb Unity Catalog.
- Quan busques **reduir l'orquestració manual** i assegurar execucions fiables en entorns distribuïts.

Evita Lakeflow Jobs si només tens una **tasca senzilla** (p. ex. una ingesta puntual), ja que en aquest cas és més eficient utilitzar **Ingestion Pipeline o ETL Pipeline**.



Job
Orchestrate notebooks, pipelines, queries and more

Configuración y edición de trabajos de Lakeflow: Azure Databricks | Microsoft Learn

4.4

Eines pròpies de Databricks
Lakehouse Federation

Eines pròpies de Databricks

Lakehouse Federation

La federació de bases de dades és una tècnica que permet connectar via JDBC i consultar dades que es troben en la/les bases de dades dels departaments aprofitant la seva capacitat de còmput, com si formessin part de la PTD però sense la necessitat de copiar-les o moure-les físicament.

Consulta en origen: Pushdown de consultes al sistema extern (MySQL, PostgreSQ, Redshift, Oracle, etc.).

Integració amb Unity Catalog: Governança centralitzada, control d'accés i llinatge.

Sense ingesta prèvia: evita ETL complexes i redueix la duplicació de dades.

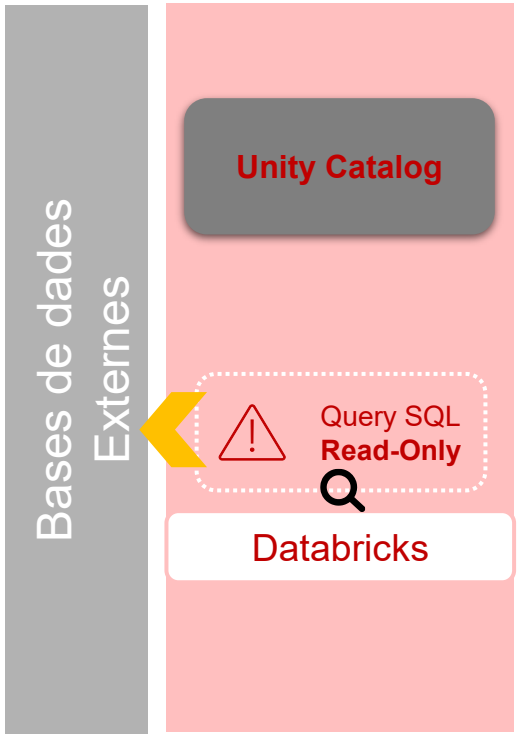
BBDD suportades per Databricks
https://docs.databricks.com/aws/en/query-federation/

CONTRES	**PROS**
El rendiment dependrà de la infraestructura origen.	Governança unificada amb Unity Catalog.
No apte per grans volums de dades ni baixa latència. En aquests casos millor ingestar les dades. Si es fan joins grans **pot impactar al rendiment.**	Permet consultar les dades **sense la necessitat de copiar-les o moure-les** físicament.
Només suporta lectura no escriptura. No pot realitzar a l'origen operacions de transformació complexes.	Ideal per reporting i prototips doncs es mecanisme àgil per accedir a les dades i menys costos que d'altres opcions.
No es possible utilitzar les UDFs de Unity Catalog	

Eines pròpies de Databricks

Lakehouse Federation

Consell:
Utilitza **Lakehouse Federation** quan necessitis consultar dades en sistemes externs sense moure-les a Databricks, per obtenir informació ràpida (real-time) en informes o anàlisi exploratori.



5

Eines disponibles a la PTD integrades amb Databricks

5.1

Eines disponibles a la PTD integrades amb Databricks

Fivetran

Eines disponibles a la PTD integrades amb Databricks

Fivetran

Fivetran és una eina SaaS per **ingestió automatitzada** que connecta múltiples fonts (BBDD, SaaS, APIs, fitxers) i carrega dades a Databricks. Ideal per ingestes **ràpides i simples** sobretot **per equips d'analítica i Data Science**. No recomanable per processos productius perquè no integra amb Git ni compleix requisits estrictes de seguretat:

- El connector per a Databricks en el moment de redacció d'aquesta guia es una versió Beta.
- No esta integrat amb el Git.
- Degut a això, el model SaaS no és compatible amb els requeriments de seguretat de la PTD.

És adequat per a **ingestes ràpides i simplificades de dades** per equips de Data Science o d'analítica.

Fivetran s'encarrega **d'extreure automàticament** aquestes dades des de les fonts connectades (bases de dades, aplicacions, APIs, fitxers).

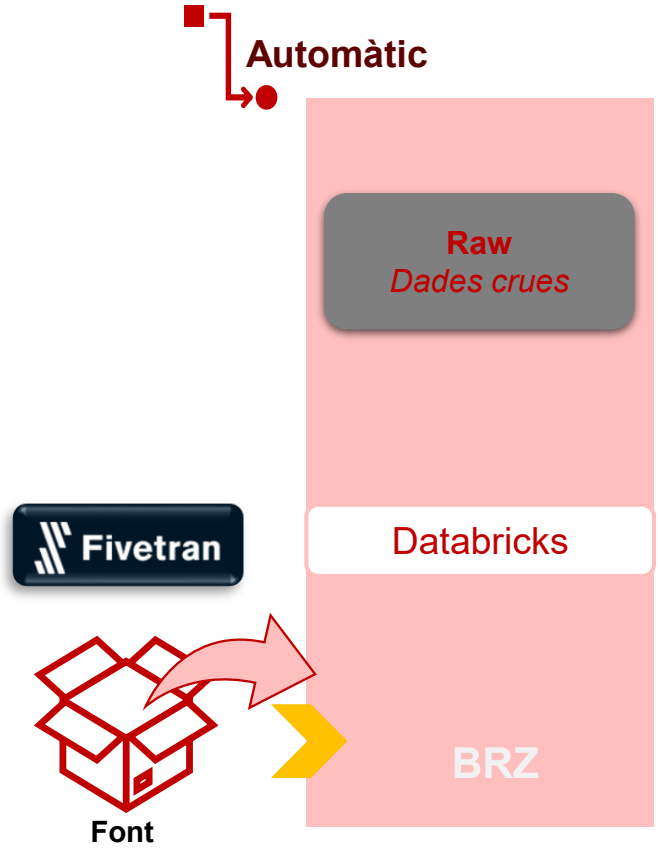
Llistat de connectors: https://fivetran.com/docs/connectors
Disponible per a multi-proveïdor a partir d'octubre 2025

CONTRES	PROS
No recomanable per a processos productius doncs no esta integrat amb GIT i no compleix amb el req. de seguretat de la PTD.	Suporta connectors per a moltes BBDD, APIs de SaaS (ex. Salesforce), i altres sistemes
Cost alt en grans volums: el preu es basa en nombre de connectors i volum de dades.	**Ideal per ingestes ràpides i sense codi**, especialment per equips d'analítica i DS.
Menys flexibilitat per lògica complexa: no substitueix pipelines avançats com DLT.	**Redueix temps de desenvolupament** i accelerar proves de concepte.
Limitacions en personalització: transformacions avançades requereixen eines extra.	**Automatització completa**: sense scripts ETL manuals.

Eines disponibles a la PTD integrades amb Databricks

Fivetran

Consell:
Utilitza Fivetran quan necessitis **ingestes ràpides i automatitzades** des de múltiples fonts (BBDD, SaaS, APIs) cap a Databricks, especialment per **carregar dades crues a la capa Bronze** (BRZ) **sense mantenir scripts ETL**. Fivetran s'encarrega d'extraure automàticament les dades i dipositar-les en Databricks, permetent que després s'apliquin altres eines per garantir qualitat i transformacions avançades.



5.2

Eines disponibles a la PTD integrades amb Databricks

Denodo

Eines disponibles a la PTD integrades amb Databricks

Denodo

Denodo és una plataforma de virtualització de dades que permet accedir a dades des de múltiples fonts sense moure-les, creant una capa d'abstracció que facilita consultes unificades via JDBC. És ideal quan les dades no necessiten aplicar qualitat i es vol evitar processos ETL complexos. També permet crear vistes virtualitzades i integrar fonts heterogènies (BBDD, SaaS, fitxers).

Característiques clau:

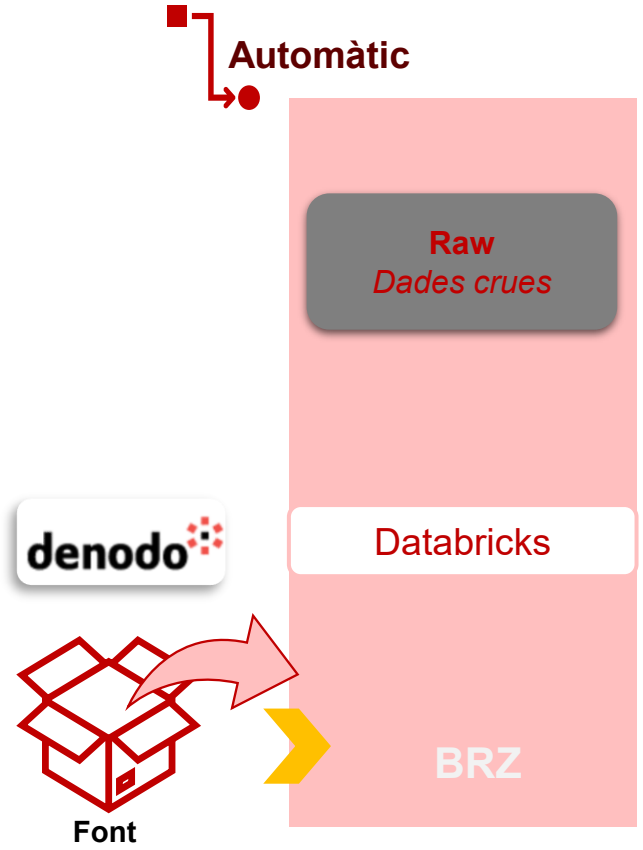
- Virtualització en temps real: accés sense replicar dades.
- Suport per múltiples connectors (incloent SaaS com SharePoint).
- Creació de vistes i manipulació de dades sense ETL.
- Integració amb Databricks via JDBC per ingestió posterior si cal persistència.

CONTRES	PROS
VQL té limitacions: no suporta totes les funcions SQL (ex. rownum())	Accés en temps real a dades sense moure-les.
No emmagatzema dades de forma permanent: cal persistir-les a Delta Lake si es vol historial.	**Permet crear vistes virtualitzades** i combinar fonts heterogènies.
Cost elevat per llicències en entorns grans.	**Ideal per consultes ràpides** sense processos ETL.
No substitueix ETL avançat: per transformacions complexes cal altres eines.	Compatibilitat amb molts connectors (ex. SharePoint, SaaS).
	Redueix duplicació i costos d'emmagatzematge.

Eines disponibles a la PTD integrades amb Databricks

Denodo

Consell:
Utilitza Denodo quan necessitis **accedir a dades de múltiples fonts sense moure-les**, gràcies a la seva capa d'abstracció que virtualitza l'origen i permet consultes via JDBC. És ideal per a escenaris on les dades **no requereixen aplicar qualitat immediata i es vol evitar processos ETL complexos**. Pots combinar Denodo amb Databricks per persistir dades a la capa Bronze (BRZ) si cal historial o transformacions avançades.



5.3

Eines disponibles a la PTD integrades amb Databricks

Event Hubs

Eines disponibles a la PTD integrades amb Databricks

Event Hubs

Event Hubs és un servei d'**ingesta massiva d'esdeveniments en temps real** d'Azure, ideal per capturar fluxos de dades (telemetria, IoT, logs, etc) i processar-los amb Databricks mitjançant Spark Structured Streaming. Aquesta integració és adequada per ingestes **contínues i escalables**, especialment per equips de Data Engineering i Data Science que necessiten dades en temps real.

Característiques clau:

- **Connexió directa** amb connector oficial d'Event Hubs per Spark i/o amb connector Kafka.
- Lectura en streaming i **escriptura en Delta Lake** (capa Bronze).
- Suport per **checkpointing i tolerància a fallades**.

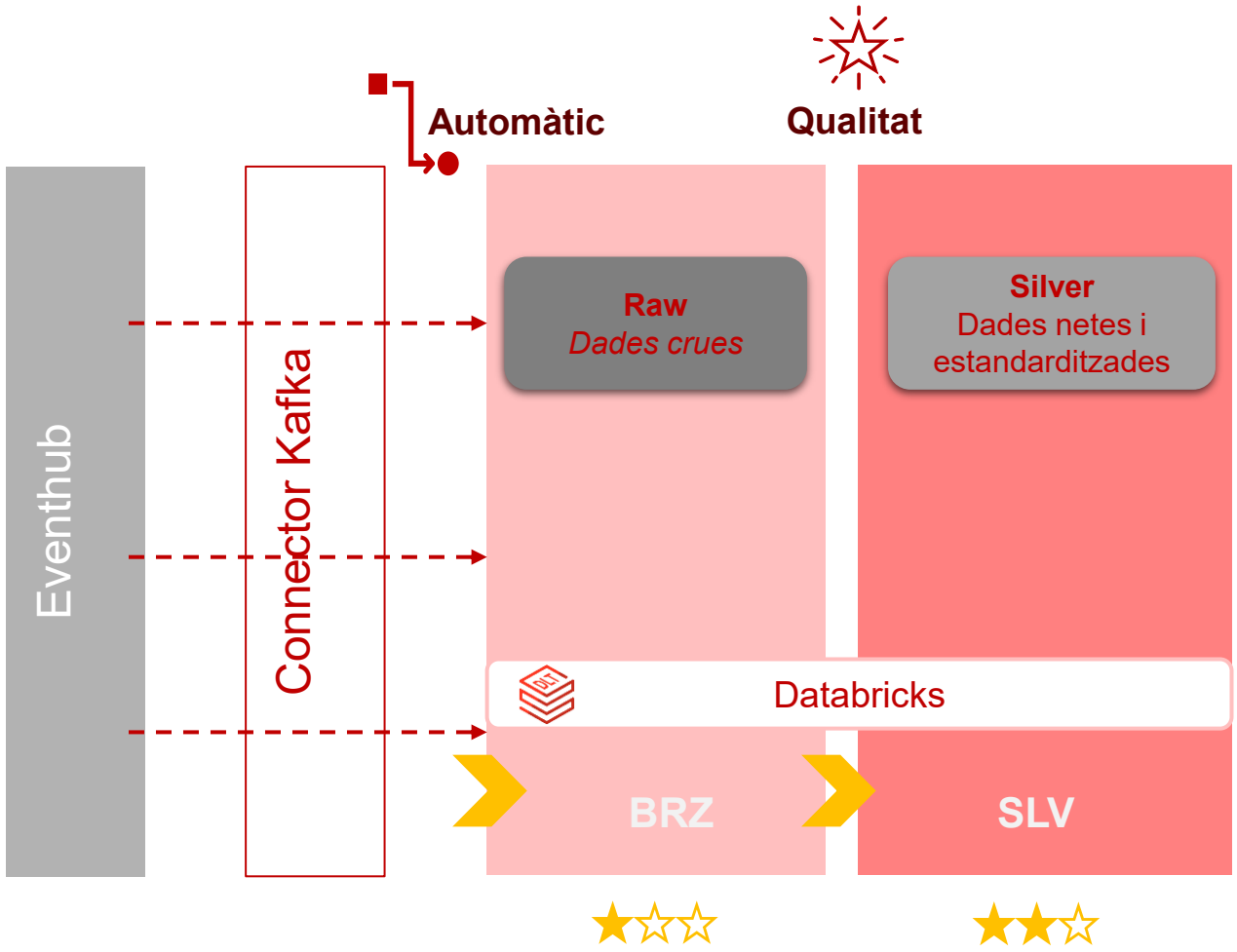
https://community.databricks.com/t5/technical-blog/high-performance-streaming-from-azure-event-hubs-using-apache/ba-p/95297

CONTRES	PROS
Requereix una configuració via YAML i un desplegament per a cada cas d'ús.	Ideal per ingestes en temps real i escalables.
El streaming pot incrementar costos si el flux és constant i massiu.	Integració nativa amb Databricks i Delta Lake.
Menys adequat per fluxos purament batch. Encara que es podria mitjançant Event Hubs Capture o Jobs Databricks planificats.	Compatible amb pipelines Bronze → Silver → Gold.
Menys adequat per a processos d'ETL complexos tot i que les transformacions avançades es poden fer manualment amb Spark.	Redueix temps de latència per aplicacions analítiques i ML.

Eines disponibles a la PTD integrades amb Databricks

Event Hubs

Consell:
Utilitza Event Hubs quan necessitis **ingestió en temps real i alt volum d'esdeveniments**. És especialment adequat si tens fluxos continus de dades com telemetria IoT, logs d'aplicacions, clics d'usuari o missatges que han de ser processats immediatament per sistemes analítics o pipelines de Streaming. Quan el teu cas d'ús requereix **baixa latència, escalabilitat i integració amb serveis d'Azure** (Databricks, Data Lake, Stream Analytics), Event Hubs és la millor opció. Quan el teu procés és purament batch, amb càrregues puntuals i sense requisits de temps real, evita Event Hubs ja que pot incrementar costos i complexitat.



6.4

Eines disponibles a la PTD integrades amb Databricks
ITQ

Eines disponibles a la PTD integrades amb Databricks

ITQ

Aquest mètode d'ingesta està **dissenyat a mida** per a la PTD per a casos d'ús amb **gran volumetria de dades** què necessiten **garantir la qualitat** de les dades i l'aplicació rigorosa de les **regles de negoci,** amb l'objectiu d'obtenir conjunts de dades **fiables i consistents.** Aquestes dades estan preparades per al seu consum en aplicacions i eines que exigeixen complir uns requisits mínims i específics de qualitat.
La informació s'emmagatzema al Delta Lake, a SFTP, a Azure amb CSV o a una altra base de dades destí del departament.

Fonts:

- SFTP o fitxer (parquet, csv, tsv, xml, json, etc.)
- Autoloader
- API
- Spark Streaming
- MongoDB
- Base de dades relacional (externa o interna)

Publicació:

- Base de dades relacional (externa o interna)
- SFTP o fitxer csv

CONTRES	PROS
Requereix corba d'aprenentatge	Per a facilitar la configuració es disposa d'un frontal web
Ingestes via API necessiten un Job a mida que descarregui la informació de l'API i la desi a volum del Azure.	ACID (Atomicitat, Consistència, Aïllament, Durabilitat). Apte per a volums grans i petits i per remediacions manuals (frontal web)
Publicacions molt complexes pot-ser necessiten d'un job a mida.	Permet aplicar regles qualitat i de negoci (ex. normalitzador adreces)
Si no cal aplicar qualitat no surt a compte pel cost que té.	Les dades RAW sempre es mantenen i tenim traçabilitat dels canvis que es van produint. A més, permet varies publicacions en paral·lel.

Eines disponibles a la PTD integrades amb Databricks

ITQ

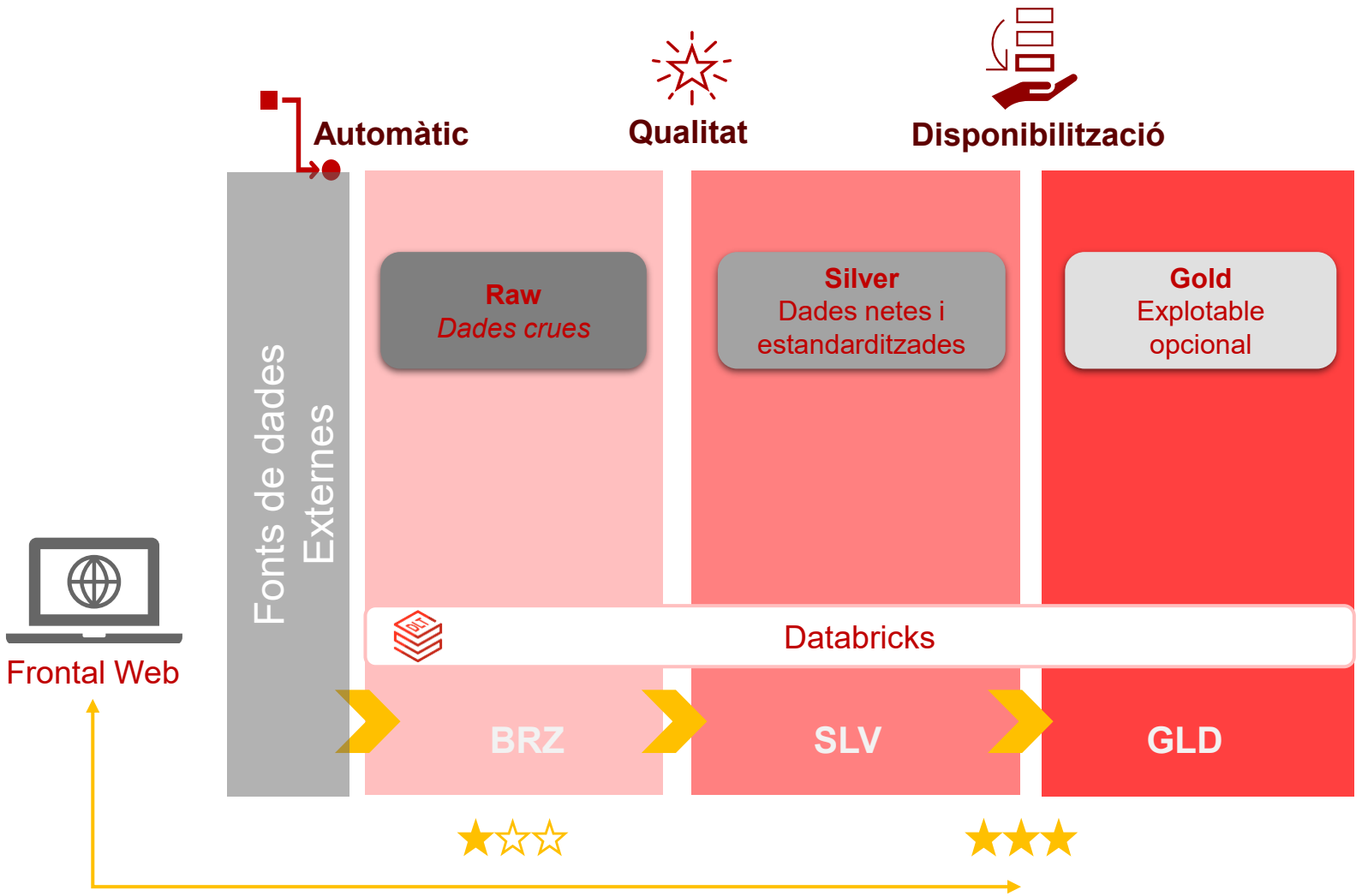
Consell:
Utilitza ITQ **quan la ingesta no és només "carregar dades", sinó que necessites assegurar qualitat i traçabilitat**. És ideal en escenaris on.

✅ **Hi ha múltiples fonts heterogènies** (SFTP, API, bases relacionals, Streaming).

✅ **Cal aplicar regles complexes** (validació, normalització, remediació).

✅ **Necessites mantenir RAW i generar capes Silver/Gold** per a explotació analítica.

✅ **Les publicacions en paral·lel i la traçabilitat dels canvis** són crítiques.



6.5

Eines disponibles a la PTD integrades amb Databricks

Framework de Plantilles

Eines disponibles a la PTD integrades amb Databricks

Framework de Plantilles

El **Framework de Plantilles** és una eina pròpia desenvolupada ad-hoc per a la PTD per simplificar el desenvolupament de determinades pipelines de dades complexes.

Destaca per tenir una estructura modular reutilitzable i **configurable mitjançant** l'ús de fitxers **YAML** destinats a aquest propòsit.

El **Framework de Plantilles** és una arquitectura modular basada en "steps" (passos) independents per a la lectura, transformació i escriptura de dades al Lakehouse, totalment parametritzables i documentats.

Pot llegir dades des d'un fitxer CSV local així com des d'un SFTP. També pot moure els fitxers que hi ha dins del SFTP i durant l'execució de la plantilla com a últim pas pot escriure el resultat del procés cap a una taula o moure-les cap a un servidor SFTP.

CONTRES	**PROS**
La font té que ser un fitxer CSV o SFTP, actualment no permet connexions a bases de dades origen i/o altre tipus de fonts origen.	Agilitat a l'hora de desenvolupar casos d'ús senzills
No permet utilitzar les UDFs pròpies d'Spark o del catàleg comú de la PTD.	Si el combinem amb un Job podem executar en Producció.
Si les funcionalitats que ofereix el framework no són suficients pel cas d'ús, cal modificar el framework.	Permet reutilitzar components entre diferents casos d'ús, evitant haver de tornar a escriure el codi.
Requereix una configuració via YAML i un desplegament per a cada cas d'ús.	

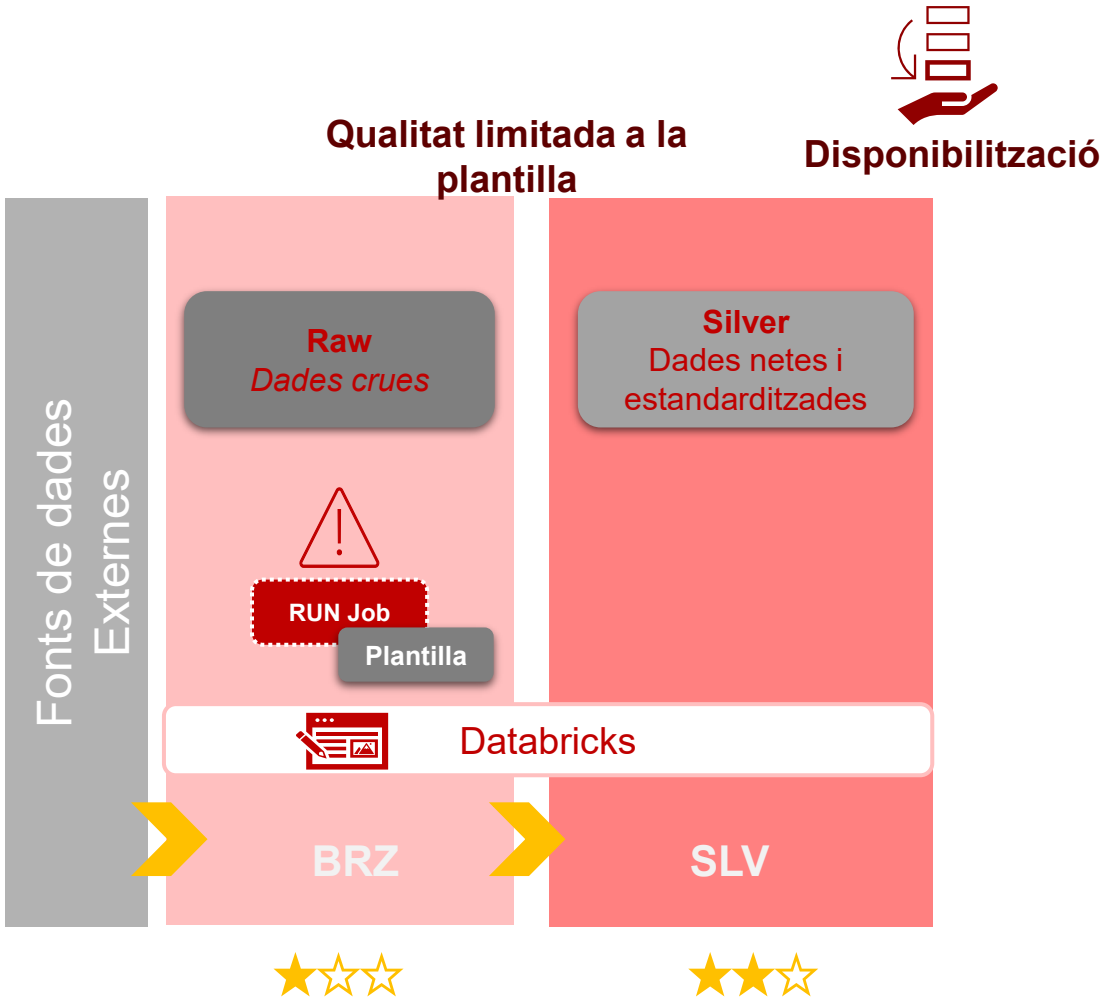
Eines disponibles a la PTD integrades amb Databricks

Framework de Plantilles

Consell:

Si volem executar-lo a producció, hem d'implementar-lo dins d'un Lakeflow Job.

La plantilla ens aporta un Framework on començar a treballar, però no cobreix totes les casuístiques per tant és per a casos d'ús molt específics.



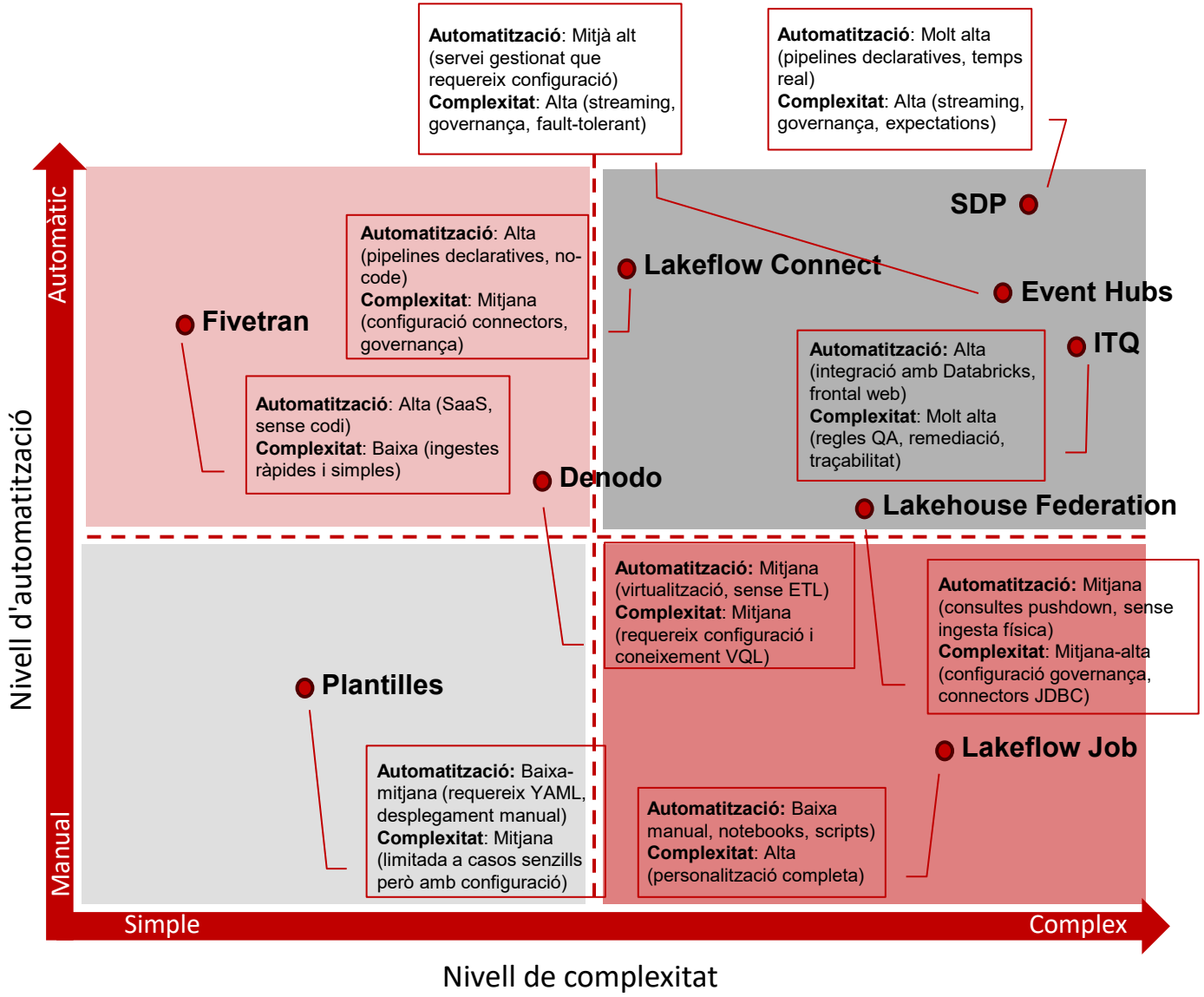
7

Criteris per la selecció del tipus d’ingesta

Criteris a avaluar

Quan seleccionem el tipus d'ingesta més adequat, hem de considerar:

- **Automatització**:
  - **Alta** → Fivetran, Lakeflow Connect, Event Hubs, SDP, ITQ
  - **Baixa** → Plantilles, Lakeflow Job
- **Complexitat del cas d'ús**:
  - **Simple** → Fivetran, Plantilles
  - **Complex** → ITQ, SDP, Event Hubs, Lakeflow Job
- **Qualitat i governança**:
  - **Necessària** → ITQ, SDP, Event Hubs
  - **Opcional** → Fivetran, Denodo
- **Fonts i connectors**:
  - **Multi-font** → Fivetran, Lakeflow Connect, Event Hubs
  - **Limitat** → Plantilles (CSV/SFTP)
- **Temps real**:
  - **Sí** → SDP, Lakeflow Connect, Event Hubs
  - **No** → Plantilles, Lakeflow Job
- **Volumetria**:
  - **Gran volum** → ITQ, Lakeflow Connect, SDP, Event Hubs
  - **Volum moderat** → Fivetran



Generalitat de Catalunya

www.gencat.cat