10b Consuming Shared Data

aws

Sources	Ingest	Transform	Query / Process	Serve	Analyse	Integrate

Sources (ETL):
- Files / Logs (semi-structured)
- Sensors and IoT (unstructured)
- RDBMS / DWH (structured)
- Business Apps (structured)
- Media (unstructured)

Federation:
- AWS Glue
- Amazon Redshift

Sharing:
- Market place
- Delta Shares

Data Intelligence Platform

Automation and Orchestration

Batch and Streaming

Data Science and AI / ML – Mosaic AI

AI/BI

Data Engineering and Processing — Spark / Photon

Data Warehousing

Data Intelligence

Operational DB

Applications

Access Control

Data and AI Governance – Unity Catalog — Catalog and Lineage

Data Management — bronze → silver → gold

Collaboration

Storage

Integrate

ID Provider
- IAM Identity Center

Governance
- Enterprise Catalog

AI Services
- Hugging Face
- Amazon Bedrock
- ...

Orchestration
- External Orchestrator

① Integrate shared metadata into Unity Catalog
② Discover shared data in Unity Catalog
③ Check access permissions and read directly from the shared object storage (for any Databricks service and client ●, since all use Spark/Photon)

Key: 3rd party / Databricks