

Databricks Cheatsheet 2026 Ready

**No Fluff quick reference
For Architects and Engineers**

Data Intelligence?

Unified Control Plane

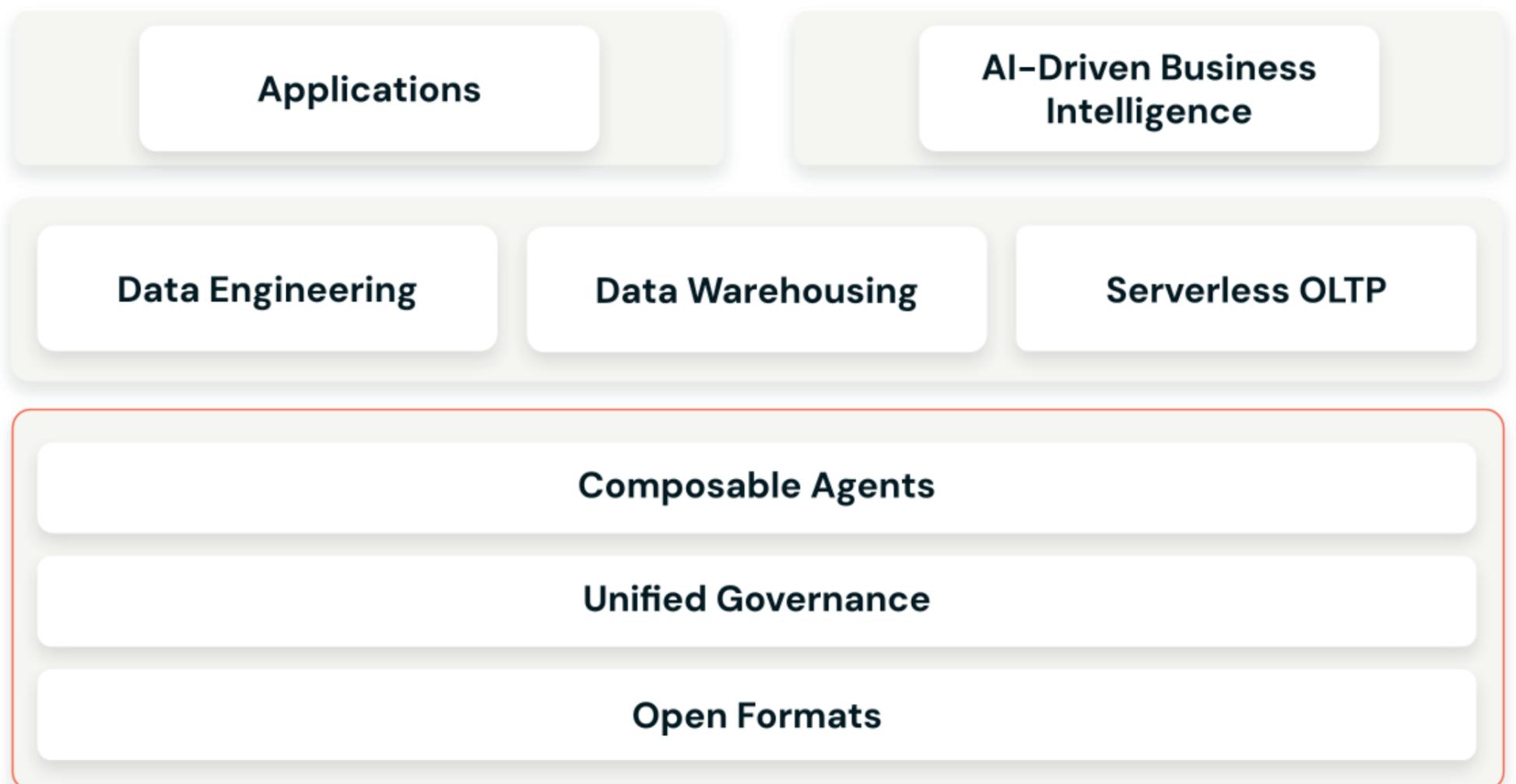
One Platform to build, deploy, share and operate data, analytics and AI across Clouds

Open Lakehouse Foundation

Keep your data in open data format, and on the same governed layer across clouds.

- ✓ Publish composable agents.
 - ✓ Build trusted tables and metrics
 - ✓ Avoid boiling the ocean!
- and more...

DATA INTELLIGENCE PLATFORM



How do you go from data to decisions with this stack?

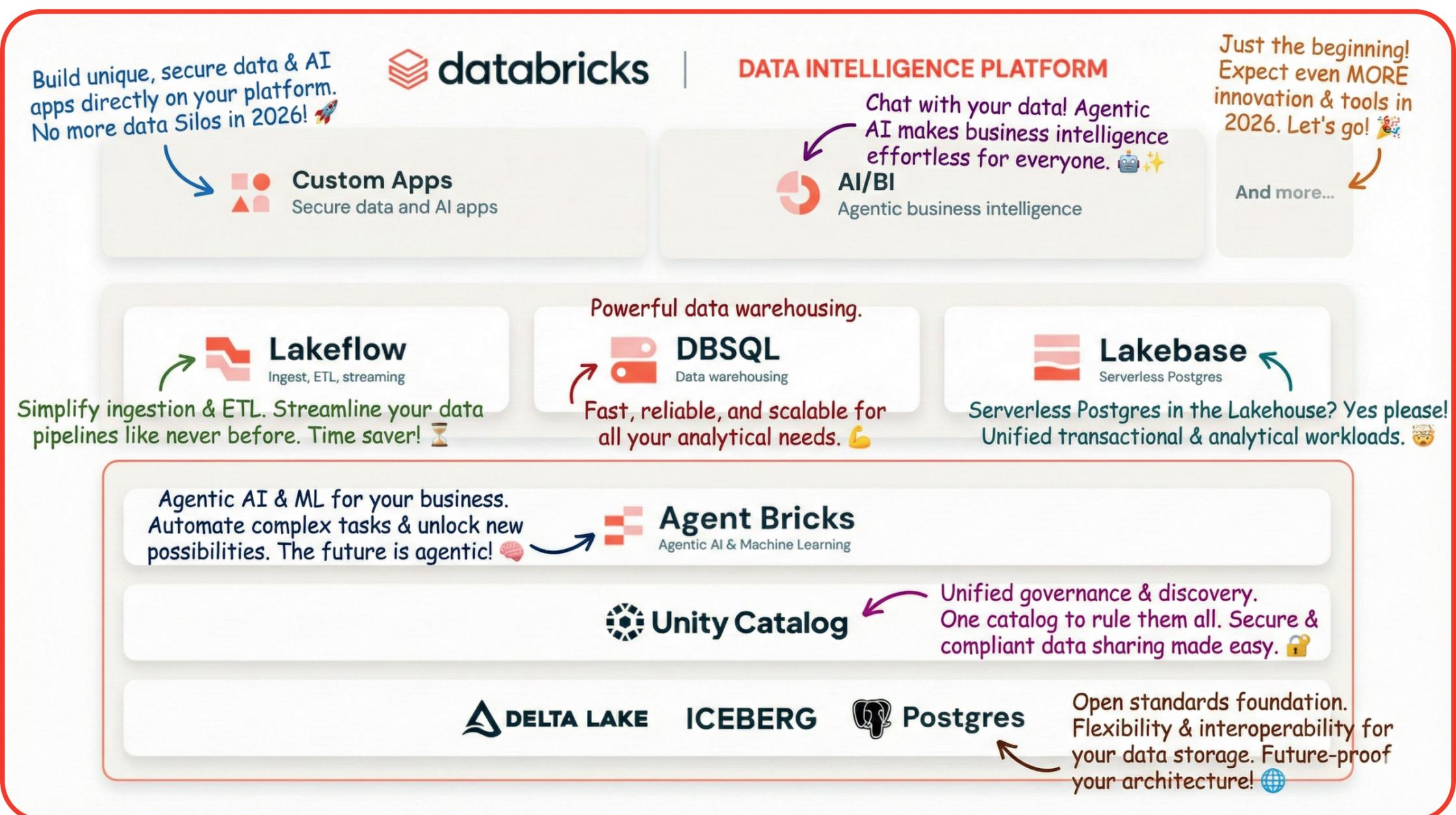
Keep reading



Databricks

Databricks unifies the full data to AI lifecycle in one place: move data in, transform it, query it with SQL, build dashboards, and deploy AI agents on top.

The key is that governance, security, and discovery sit underneath everything, so the same controls apply across every workload.





Governance and Security

Unity Catalog

Unity Catalog is Databricks' unified governance solution for centrally managing discovery, access controls, lineage, monitoring, and sharing of all data and AI assets.

Metric Views

Metric Views are governed, reusable business metric definitions in Unity Catalog that ensure consistent, SQL-native access to trusted KPIs across platforms

Access Control

Role-based and attribute-based policies that restrict or grant permissions to resources, ensuring secure and compliant data access.

Data Lineage

End-to-end tracking of data movement and transformations, providing transparency into how data is sourced, processed, and consumed.

Anomaly Detection

Automated identification of unexpected patterns or unusual behaviour in data or system performance.

Governed Tags

Standardised metadata labels are applied to data assets to enforce compliance, track sensitivity, and enable fine-grained governance.

Data Classification

Automated identification and labelling of sensitive data types to support compliance and secure data management.





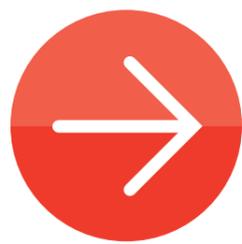
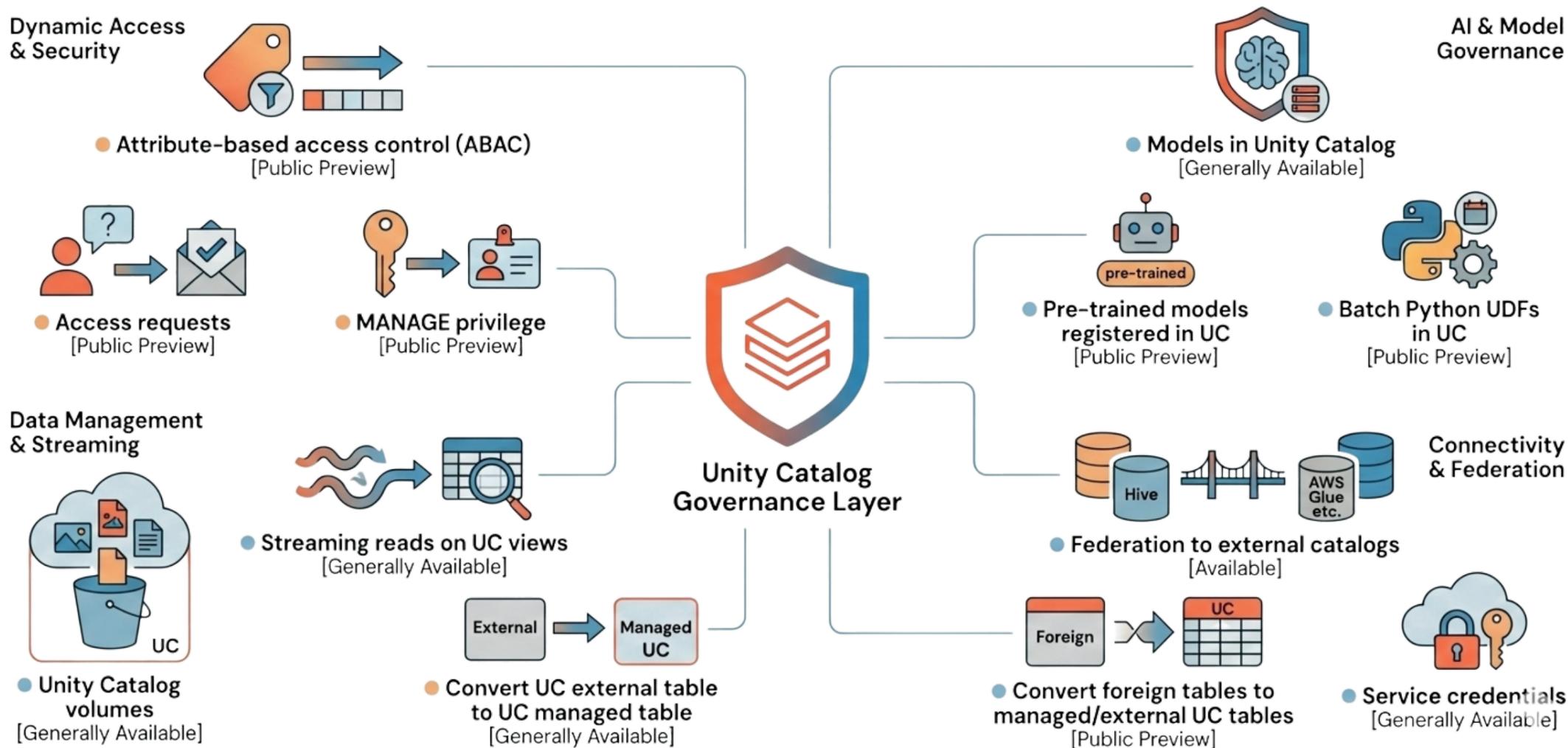
Unity Catalog



With Unity Catalog (UC), you gain centralised access control, auditing, lineage, quality monitoring, and discovery across workspaces, utilising a three-level namespace (catalog.schema.object).

There are many exciting features for 2026, out of which some of them are already in Public Preview / GA: Attribute-Based Access Control (ABAC), Scala on UC standard compute, Batch Python UDFs registered in UC and more!

Databricks Unity Catalog: Important Features up to Jan 2026





Databricks SQL



Databricks SQL runs on scalable SQL warehouses with ANSI SQL support, optimisations, and integrated tooling for queries, dashboards, alerts and semantic metrics – all governed by Unity Catalog.

It is highly integrated with Mosaic AI, which means you can utilise simple AI-powered SQL functions to offload repetitive tasks.

Some of the anticipated and highly used features in this space would be: **SQL Scripting** and procedures, Parameterisation improvements, **Query caching** enhancements and many more!

Unified with AI

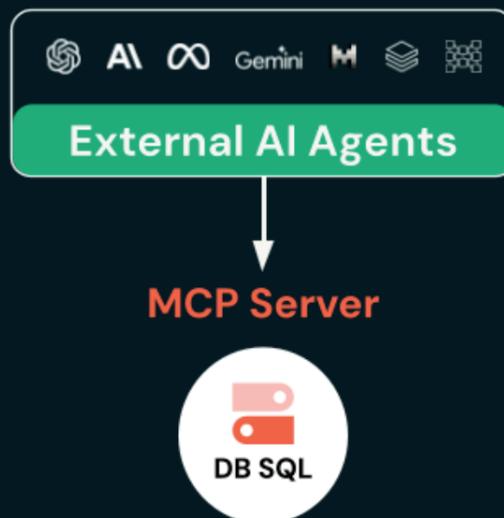
Leverage the full power of AI directly in SQL

Any LLM via SQL

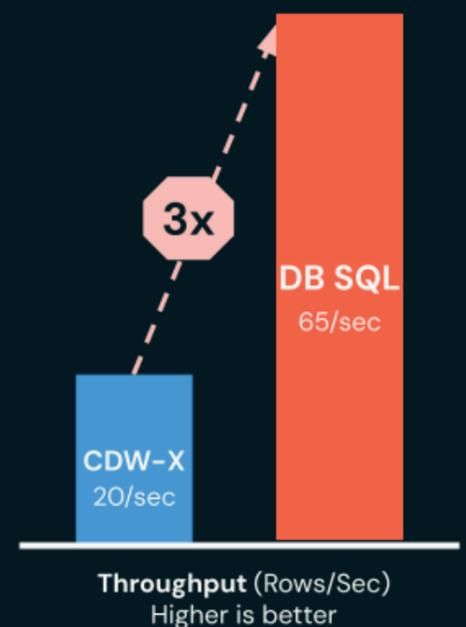
Off-the-shelf or custom models



Use DBSQL in Agents



High throughput



Data Warehousing

Databricks SQL

A collection of services that bring data warehousing capabilities and performance to users' existing data lakes.

SQL Editor

It is an Analyst-friendly user interface that blends a unified and streamlined ANSI-SQL authoring experience with multiple statement results, real-time collaboration, and enhanced Databricks Assistant integrations.

Lakehouse Federation

A collection of features that enable users and systems to run queries against multiple data sources without needing to migrate all the data to a unified system.

Query Federation

With query federation, queries are pushed down to the foreign database using JDBC APIs and executed both in Databricks and remotely on compute resources.

Catalog Federation

With catalog federation, users can directly access the foreign table in object storage using only Databricks compute.





Compute



Serverless

Serverless compute is on-demand, automatically managed compute that scales based on your workload requirements.

Classic Compute

These are provisioned compute resources that you create, configure, and manage for your workloads.

SQL Warehouse

SQL warehouses are either classic or serverless and are optimised for specific compute resources and advanced functionality.

Lakebase

A fully managed Postgres OLTP database engine that allows users to create and manage transactional databases in Databricks-managed storage, seamlessly integrate real-time OLTP workloads with your Lakehouse, and support AI/BI workloads that require lower latency queries via reverse ETL patterns.

Serverless GPU Compute

On-demand, fully managed GPU infrastructure for running ML and AI workloads without manual resource management.



Data Engineering

Lakeflow

An end-to-end data engineering solution that empowers data engineers, software developers, SQL developers, analysts, and data scientists to deliver high-quality data for downstream analytics, AI, and operational applications.

Lakeflow Connect

Lakeflow Connect simplifies data ingestion with connectors that are popular with enterprise applications, databases, cloud storage, message buses, and local files.

Spark Declarative Pipelines

Spark Declarative Pipelines is a declarative framework that reduces the complexity of building and managing efficient batch and streaming data pipelines by automatically orchestrating the execution of flows, sinks, streaming tables, and materialised views.

Lakeflow Jobs

Lakeflow jobs provide reliable orchestration and production monitoring for any data and AI workload (notebooks, pipelines, managed connectors, SQL queries, machine learning training, model deployment, and inference).

Autoloader

Autoloader incrementally and efficiently processes new data files as they arrive in cloud storage.

Lakeflow Designer

It is a no-code, AI-assisted tool for building production-grade data flows on Databricks. All designer pipelines are production-ready from day one.

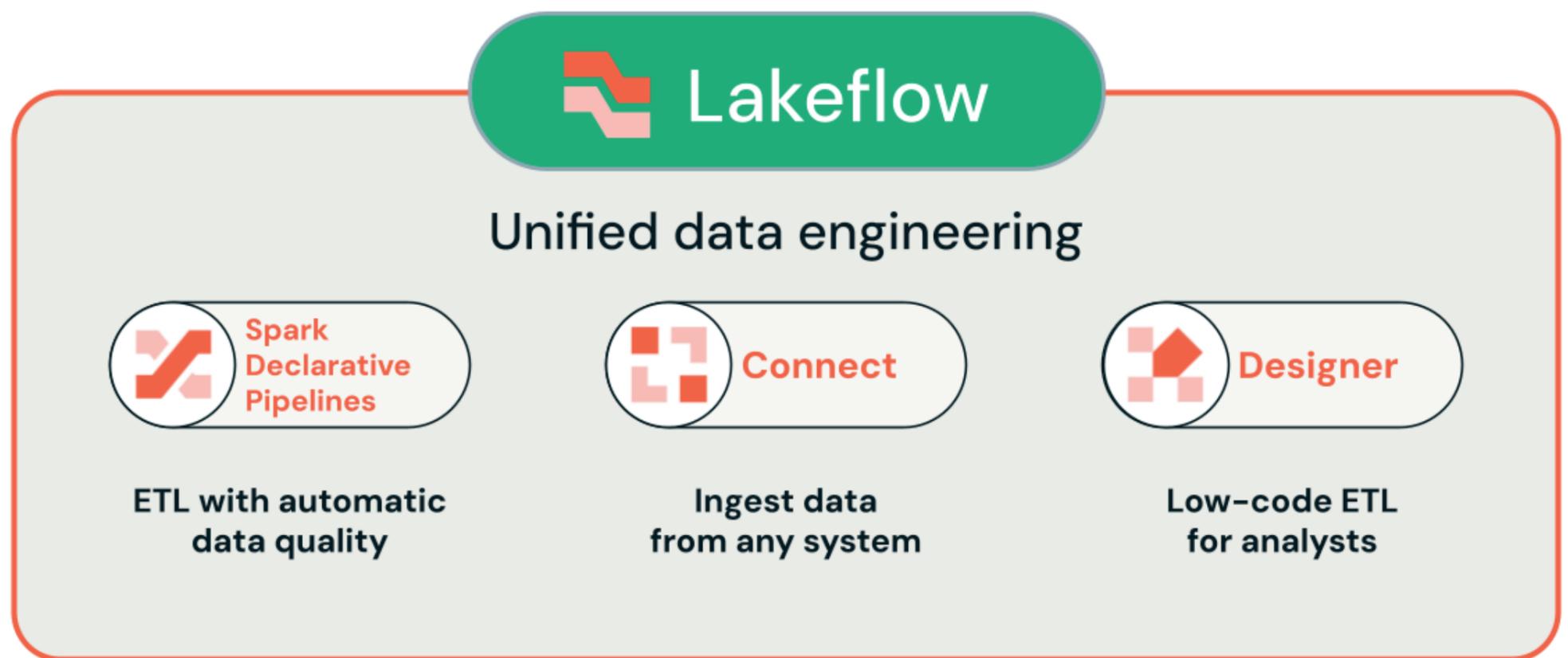


Lakeflow Ecosystem

Lakeflow unifies ingestion, transformation, and orchestration into a single, governed, serverless experience that consolidates fragmented data stacks and accelerates the delivery of reliable data for analytics and AI.

You can enable Real-Time Mode for ultra-low-latency stream processing without requiring code changes.

For modern event ingestion, **Zerobus** – a Lakeflow Connect API enables direct, high-throughput writes to the lakehouse (up to 100 MB/s) with near real-time latency (< 5s) to simplify real-time use cases.



Data Sharing

Delta Sharing

The secure data sharing platform that lets users share data and AI assets within and outside Databricks.

Clean Rooms

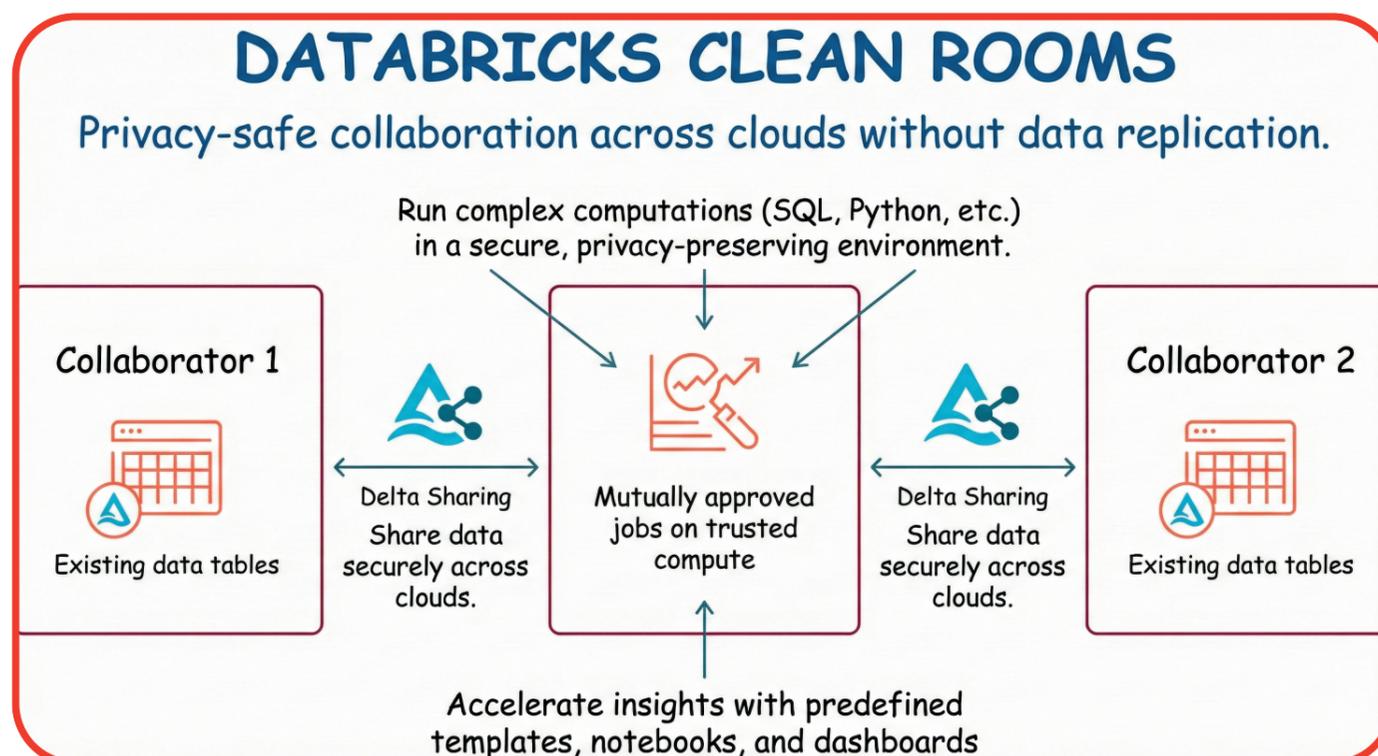
Utilises Delta Sharing and serverless compute to provide a secure and privacy-protecting environment where multiple parties can collaborate on sensitive enterprise data without requiring direct access.

Marketplace

Open forum for exchanging third-party data products.

Partner Connect

Allows users to create trial accounts with select Databricks technology partners and connect their Databricks workspace to partner solutions from the Databricks UI.





AI/BI



(Business Intelligence)

AI/BI Genie

AI/BI Genie is a Databricks feature that allows business teams to interact with their data using natural language.

AI/BI Dashboards

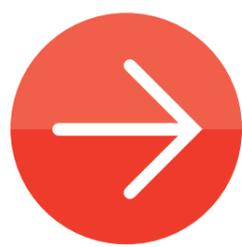
AI/BI Dashboards feature AI-assisted authoring, an enhanced visualisation library, and a streamlined configuration experience in order to transform data into data visualisations and shareable reports quickly.

Databricks Apps

Databricks Apps enables developers to build and deploy secure data and AI applications directly on the Databricks Platform, which eliminates the need for separate infrastructure.

Databricks One

A streamlined UI for workspace users focused on securely accessing published dashboards, Genie spaces, and apps without creating new content.



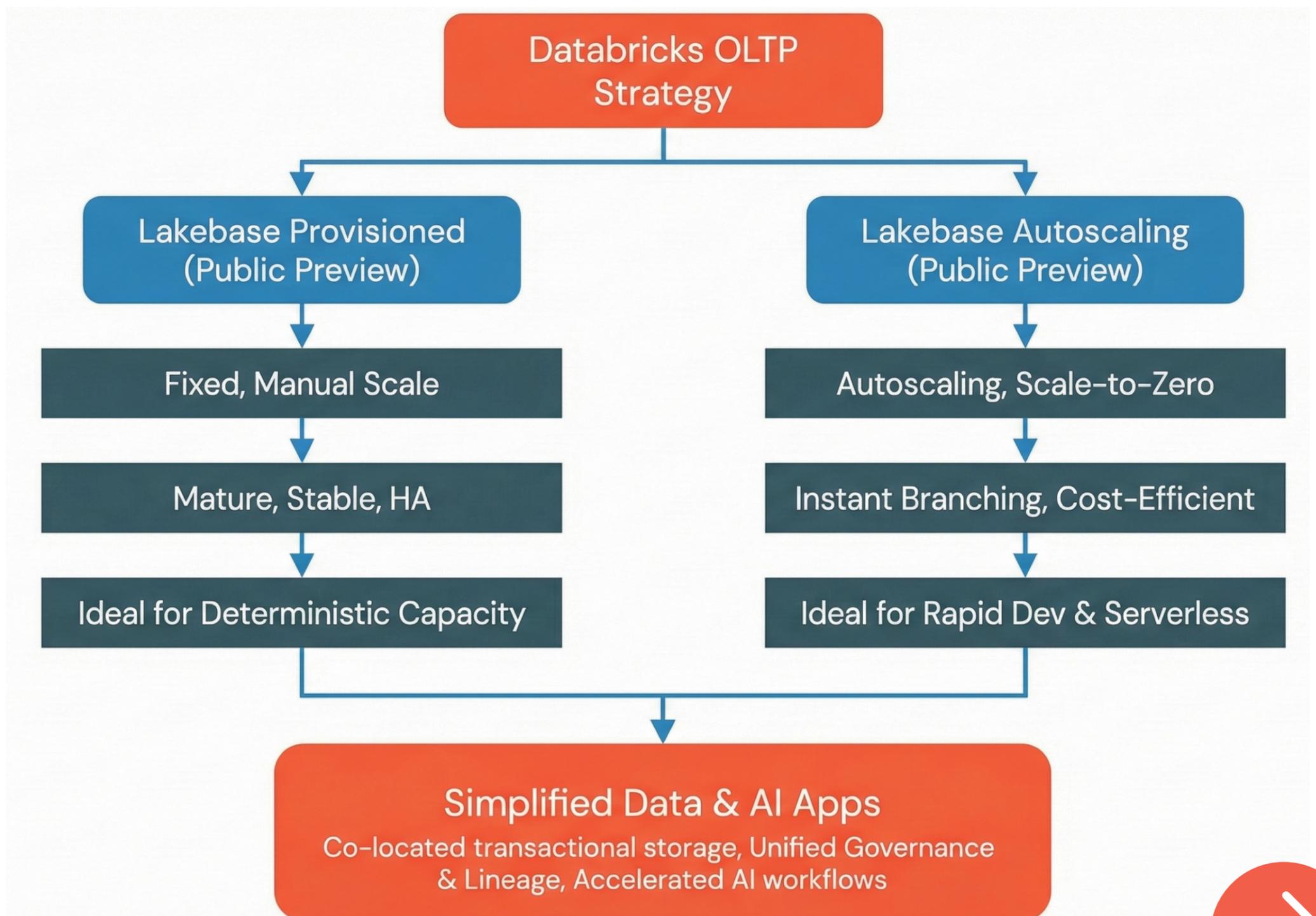


Lakebase

Lakebase is a fully managed, cloud-native PostgreSQL database that brings online transaction processing (OLTP) capabilities to the Lakehouse.

There are two variants of Lakebase at the moment:

Lakebase Provisioned
Lakebase Autoscaling

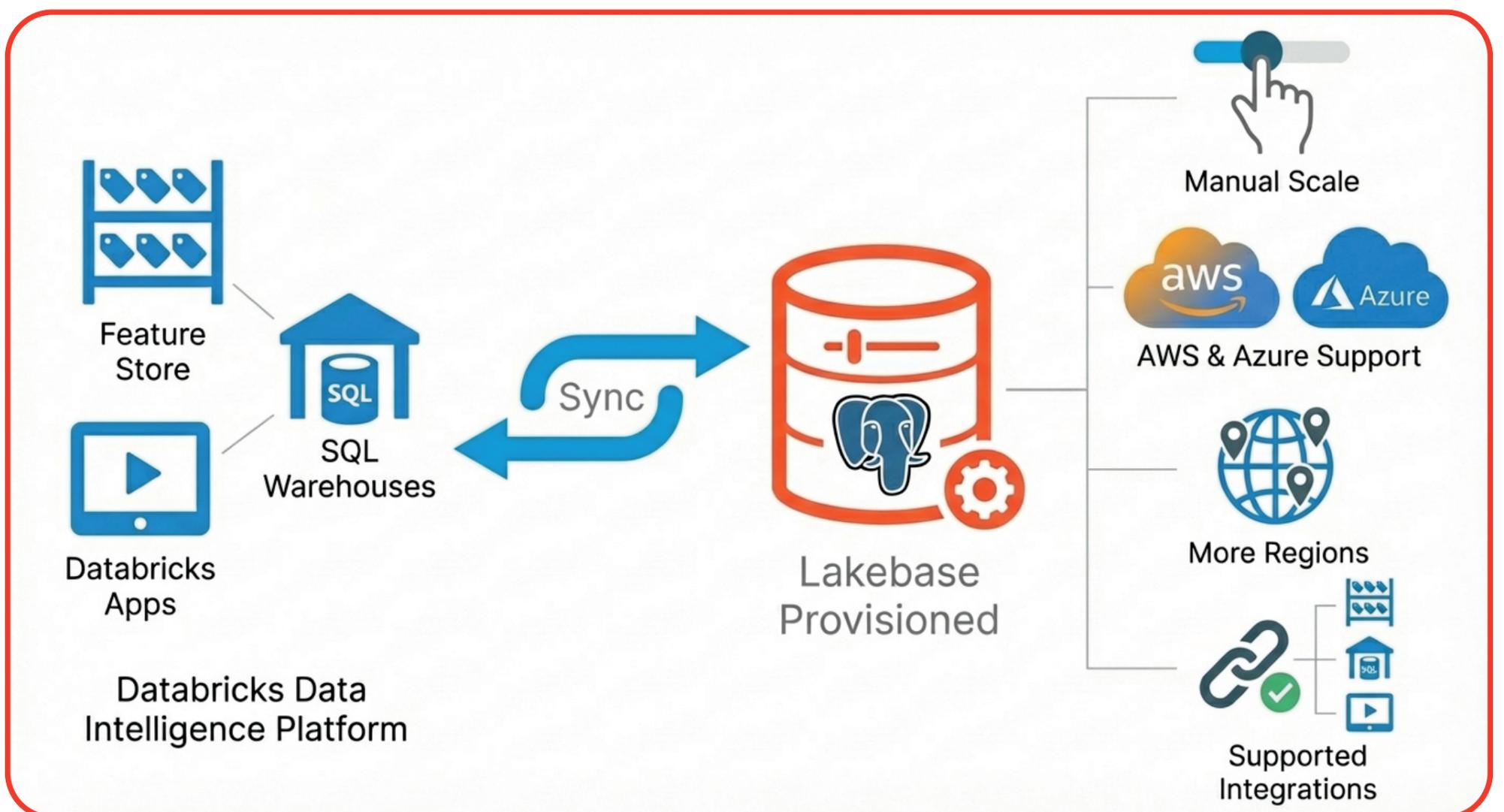




Lakebase Provisioned

Lakebase Provisioned is a fully managed PostgreSQL on AWS/Azure. If you need highly available (HA) readable secondaries or specific platform integrations today, start with **Lakebase Provisioned** today on AWS and Azure.

You can build full-stack apps with Databricks Apps backed by Lakebase and leverage native identity and fine-grained permissions.





Lakebase

Autoscaling

Lakebase Postgres Autoscaling combines the reliability and familiarity of Postgres with modern database capabilities. This enables flexible development workflows, cost-efficient operations, and rapid iteration.

Lakebase Autoscaling is the new version of Lakebase and is progressing toward feature parity with Lakebase Provisioned.

FEATURES OF LAKEBASE AUTOSCALING

Auto Scaling

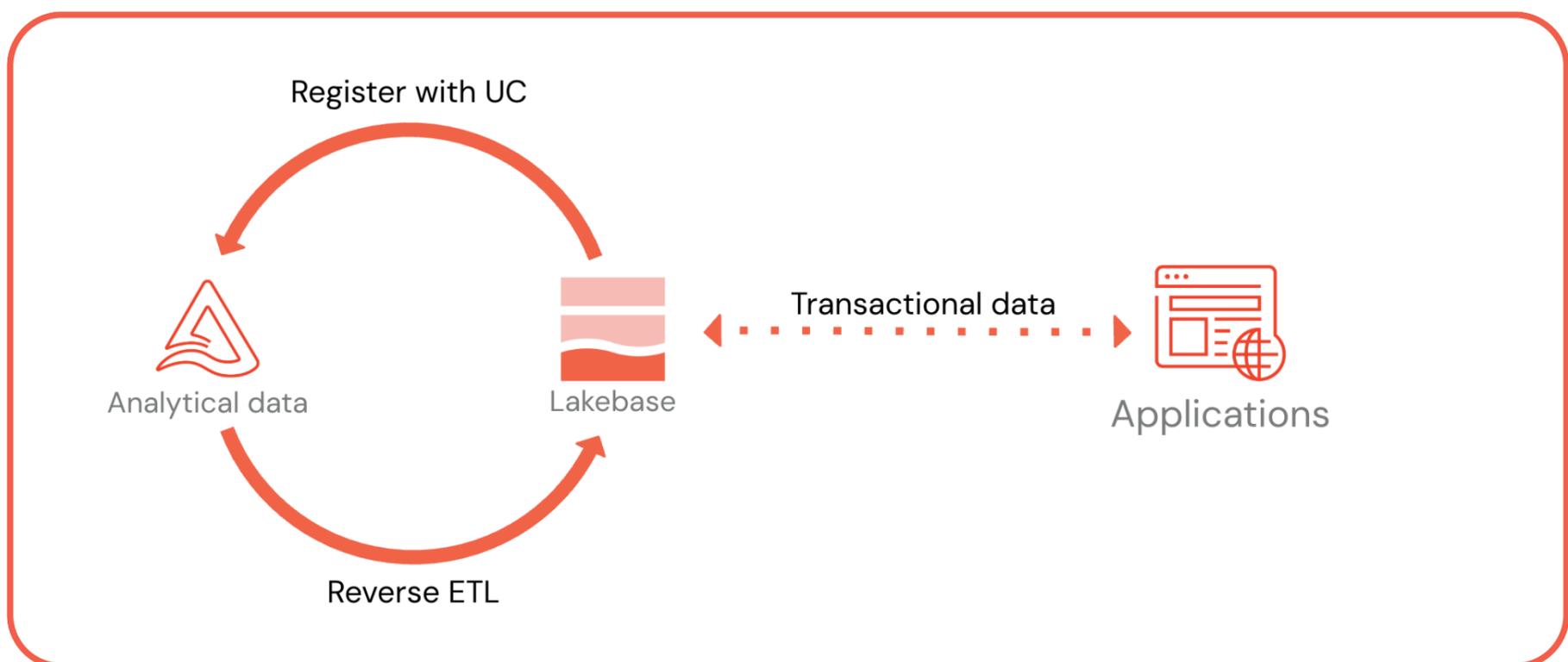
Branching

Scale-to-zero

Instant Restore

Data API

Read Replicas





Machine Learning



MLflow 3.0

An open source tool that simplifies and manages the complete machine learning lifecycle, including experiment tracking, model management, and deployment, on Databricks and beyond.

AutoML

Automated machine learning capabilities that rapidly build, select, and optimise models without extensive coding.

Feature Store

A centralised repository for storing, sharing, and governing machine learning features to streamline model development and reuse.

Machine Learning Runtime

An optimised, up-to-date runtime environment tailored for scalable machine learning and AI workloads on Databricks.





Mosaic AI

A unified suite of generative AI tools for building, deploying, and customising enterprise AI solutions on Databricks.

Model Serving

Scalable infrastructure for deploying and managing machine learning models as real-time REST APIs.

Foundation Model Serving

Managed deployment of large-scale foundation models (like LLMs) on Databricks for rapid generative AI integration.

Vector Search

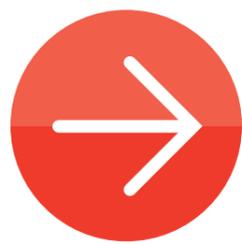
A high-performance engine for storing and searching features and embedding vectors to enable semantic retrieval and AI-powered search.

Agent Framework

A toolkit for building, orchestrating, and managing AI agents that automate tasks across enterprise workflows.

Agent Evaluation

Automated tools and dashboards for benchmarking and improving the performance and reliability of AI agents.





Gateway

A centralised interface that manages secure, unified access to deployed AI and ML models and agents within an organisation.

AI Playground

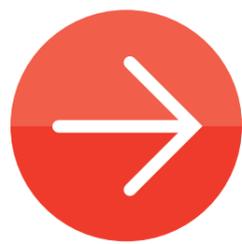
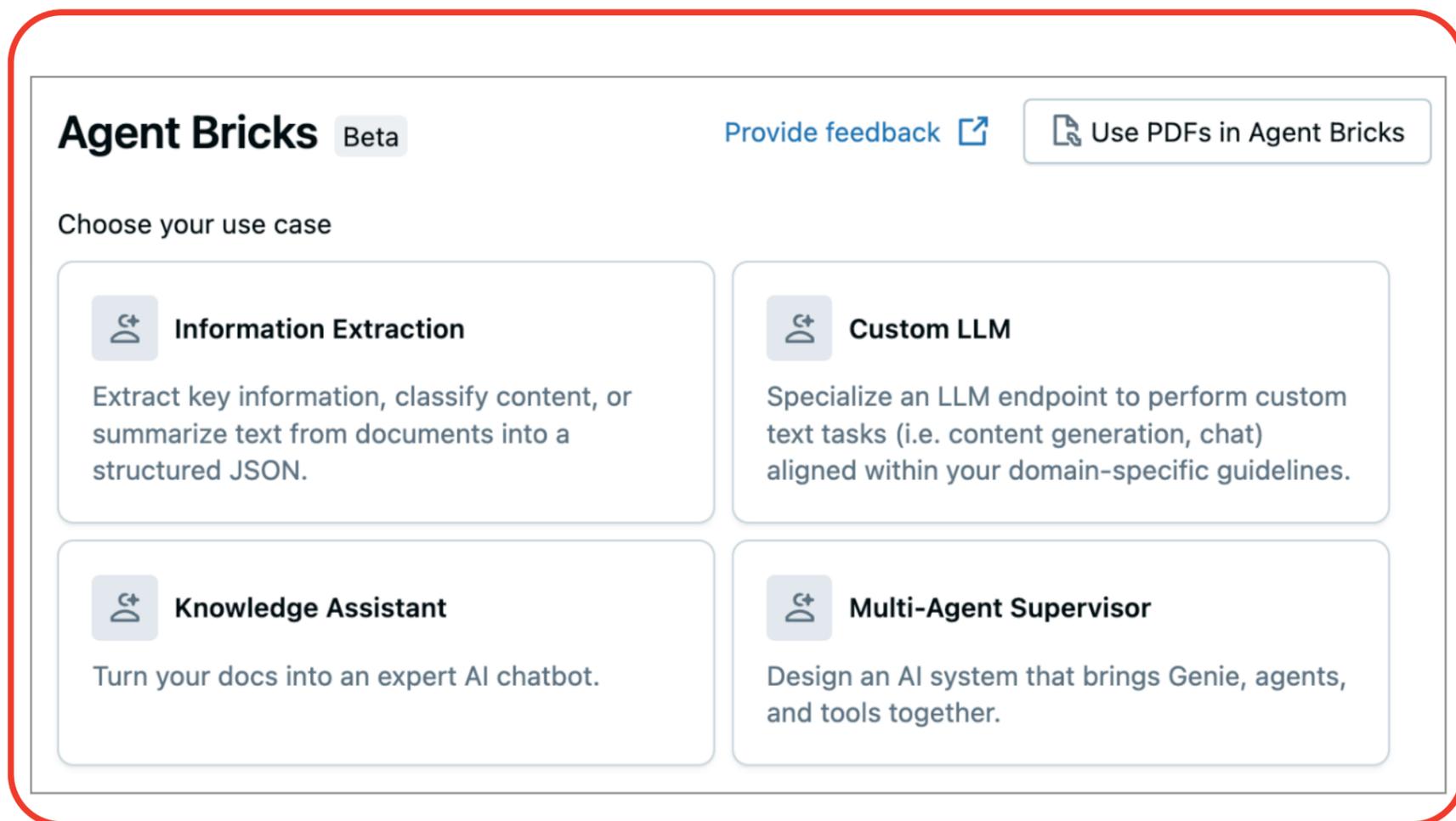
An interactive environment to experiment with models, prompts, and AI components for rapid prototyping and learning.

Agent Bricks

Reusable, composable building blocks for quickly creating and deploying AI agent workflows in Databricks.

Inference Tables

Structured tables capturing inputs, outputs, and metadata from model predictions to enable monitoring and analysis at scale.



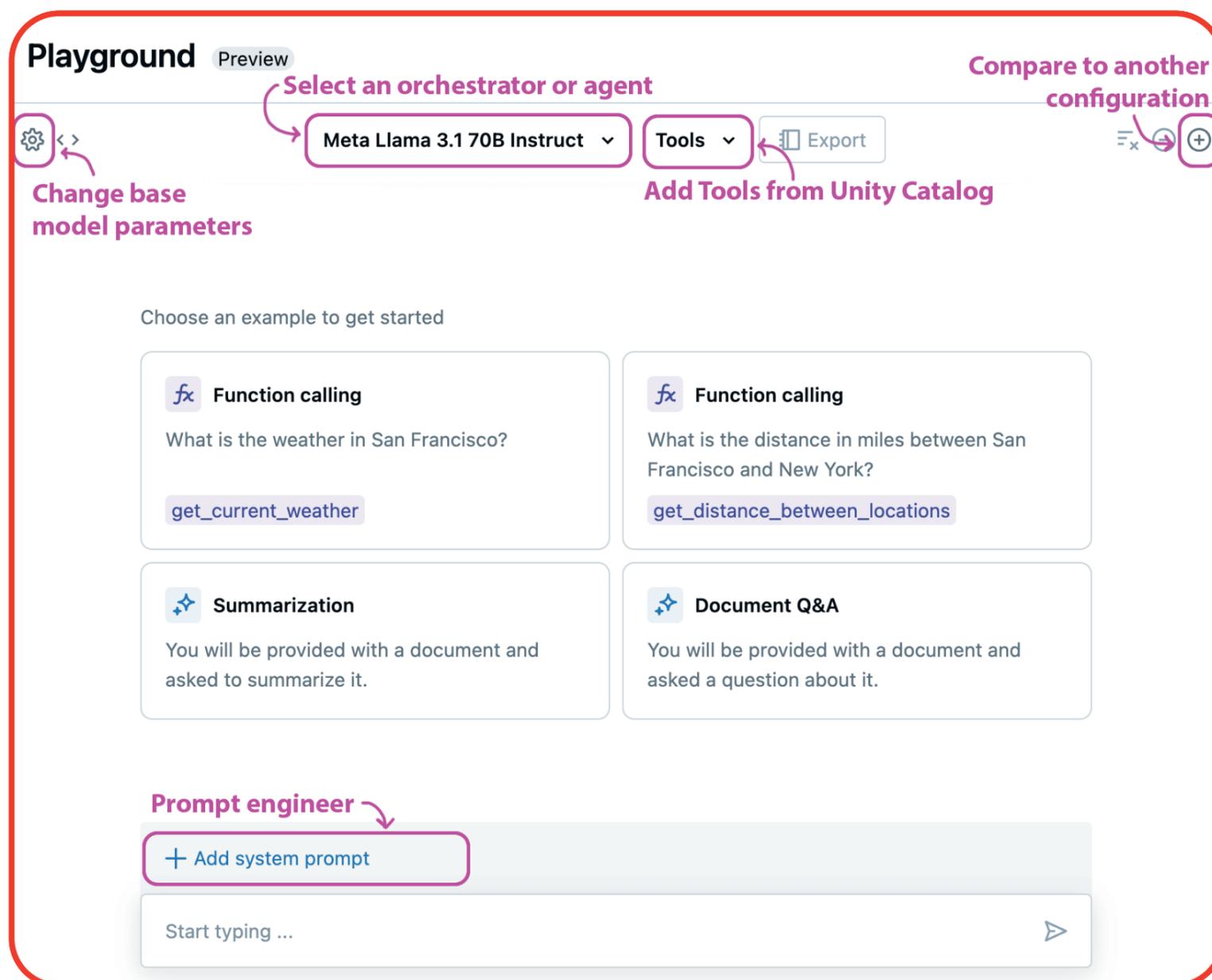


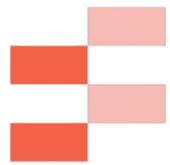
Mosaic AI



Mosaic AI is the end-to-end AI and ML suite on the Data Intelligence Platform, designed to help teams build, govern, evaluate, and deploy production-quality AI agents and models across generative and classical AI workloads.

The **AI Playground** is one of the offerings in the Mosaic AI pillar. It is a chat environment for LLMs where you can test prompts, tune parameters, and compare multiple model endpoints side by side to quickly “vibe-check” models and behaviours.





Agent Bricks

This product is offered for building and optimising **domain-specific AI agent systems**, currently in Beta, with access controlled from Workspace Previews. It simplifies agent implementation and **auto-evaluates/optimises** for quality and cost using Mosaic AI research-backed techniques.

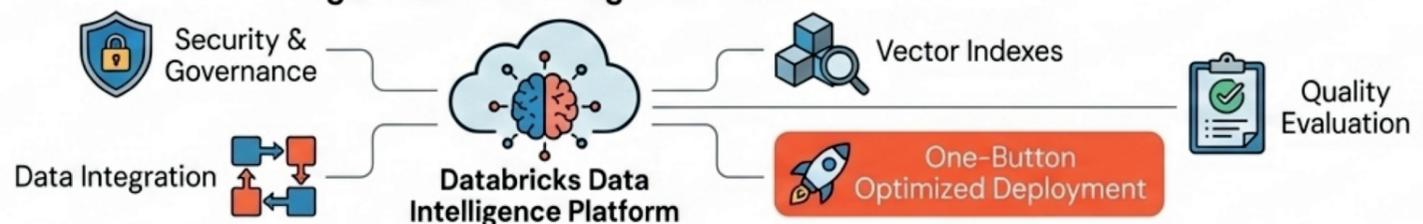
Four bricks  are currently available:

- ✓ Information Extraction
- ✓ Custom LLMs
- ✓ Knowledge Assistant
- ✓ Multi-Agent Supervisor

Agent Bricks: How it Works & Supported Use Cases

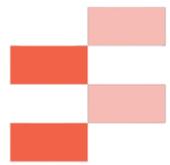


Integrated Data Intelligence Platform



Supported Use Cases





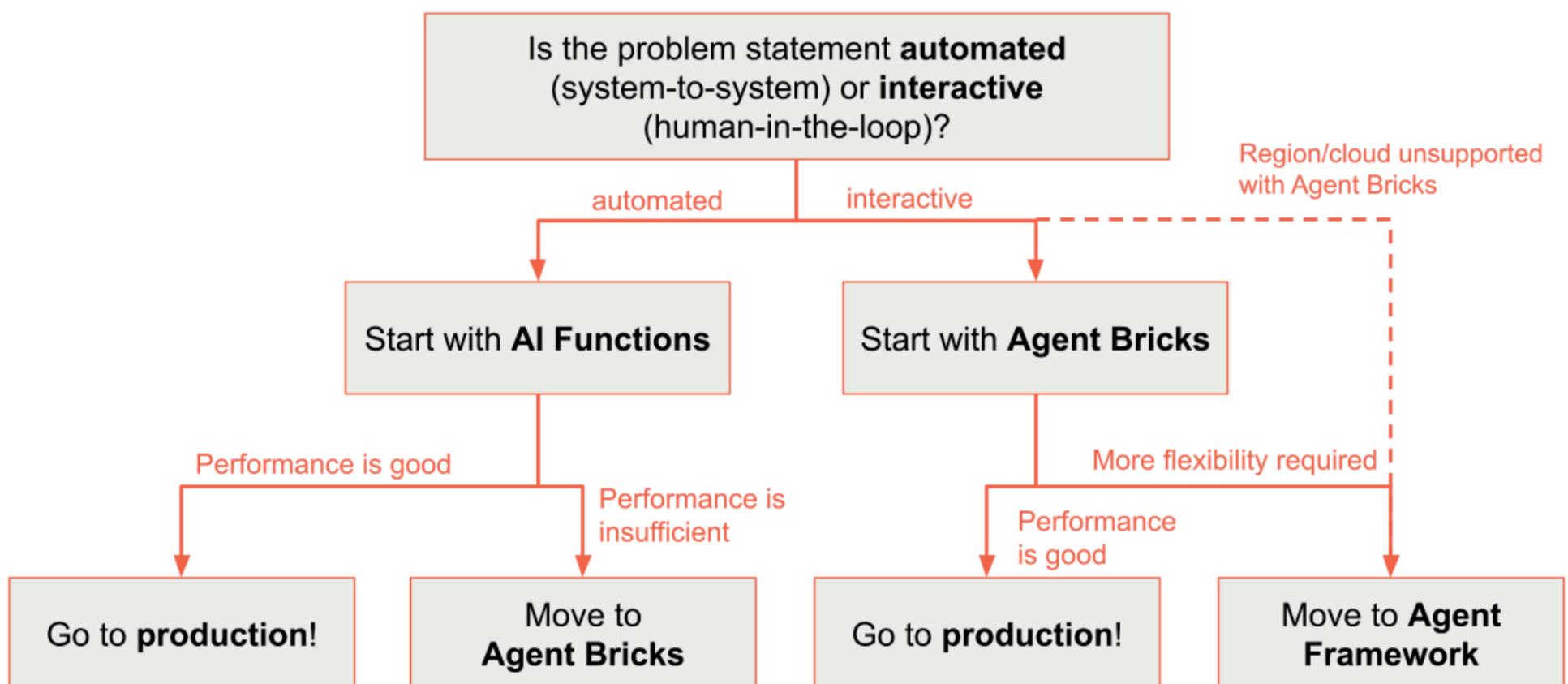
Agent Bricks

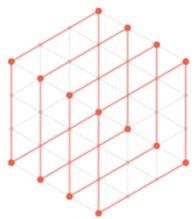
Agent Bricks helps you take AI ideas to production faster by packaging the complex parts of agent delivery into reusable building blocks. If you are unsure when to use AI Functions vs. Agent Bricks vs. the Agent Framework, this slide gives you a simple decision path.

Some common examples where Agent Bricks shines (not an exhaustive list) are below:

- ✓ Vendor name matching and deduplication using the **Information Extraction** brick
- ✓ Email and document classification using the **Information Extraction** brick
- ✓ Contract Q&A plus clause extraction using a **Custom LLM Agent**

Guide to choose the right tool





Databricks Apps

You can build and deploy secure data and AI applications directly on Databricks, hosted on serverless and integrated with Unity Catalog, Databricks SQL, Model Serving, Lakeflow Jobs, and OAuth/service principals for auth.

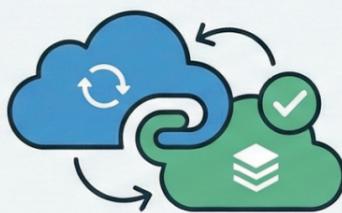
Quick tip 💡: Use App runtime logs (stdout/stderr via App Logs tab or `.../logz` URL) for quick debugging

DATABRICKS APPS: ENTERPRISE ADVANTAGES



SECURE GOVERNANCE & AUTHENTICATION

Unity Catalog Integrated & User Scoping



SEAMLESS SERVICE ACCESS

Automated SDK Auth & Service Principal



RICH NETWORK CONNECTIVITY

Serverless NCC, On-Prem & External Access



USAGE-BASED PRICING

Pay-as-you-go Model



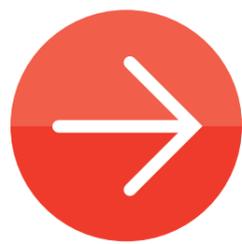
FULLY ISOLATED & SCALABLE

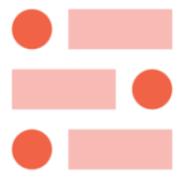
Default Isolation & Vertical Scaling



EXTENSIVE DEV SUPPORT

Large Package Support, Detailed Docs, UI Deployment





Productivity

Databricks Notebooks

Primary tool for creating data science and machine learning workflows on Databricks. They provide real-time coauthoring in multiple languages, automatic versioning, and built-in data visualisations for developing code and presenting results.

Databricks Assistant

An AI-powered pair programmer and support agent that helps you generate, optimise, and troubleshoot code, queries, and dashboards to accelerate your work. The (modes: **Chat**, **Edit**, **Agent**) across Notebooks and the SQL Editor to generate, optimise, explain, and fix code/queries, plus inline shortcuts and autocomplete support.

Databricks Data Science Research Agent

It is a powerful capability in the Databricks Assistant's Agent Mode that transforms the Assistant into an intelligent companion, capable of automating entire multi-step data science workflows in Databricks Notebooks and the SQL Editor.





Data Storage Layer



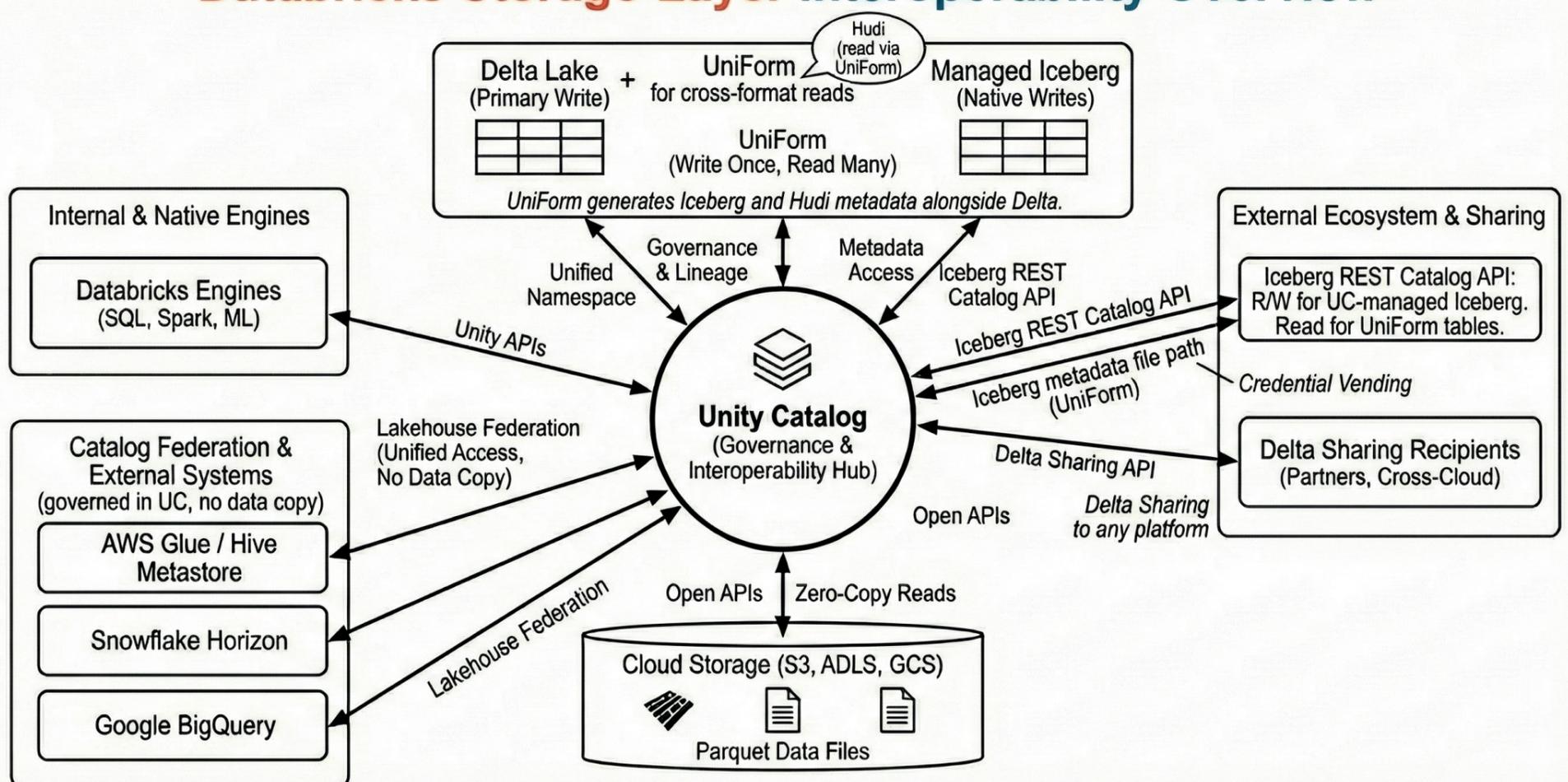
Delta

Delta is an open source table format that extends Parquet data files with a file-based transaction log for ACID transactions and scalable metadata handling.

Apache Iceberg™

Iceberg is an open source table format for analytics workloads, supporting features like schema evolution, time travel, and hidden partitioning, along with an abstraction layer that enables ACID transactions on data stored in object storage.

Databricks Storage-Layer Interoperability Overview





Monitoring

System Tables

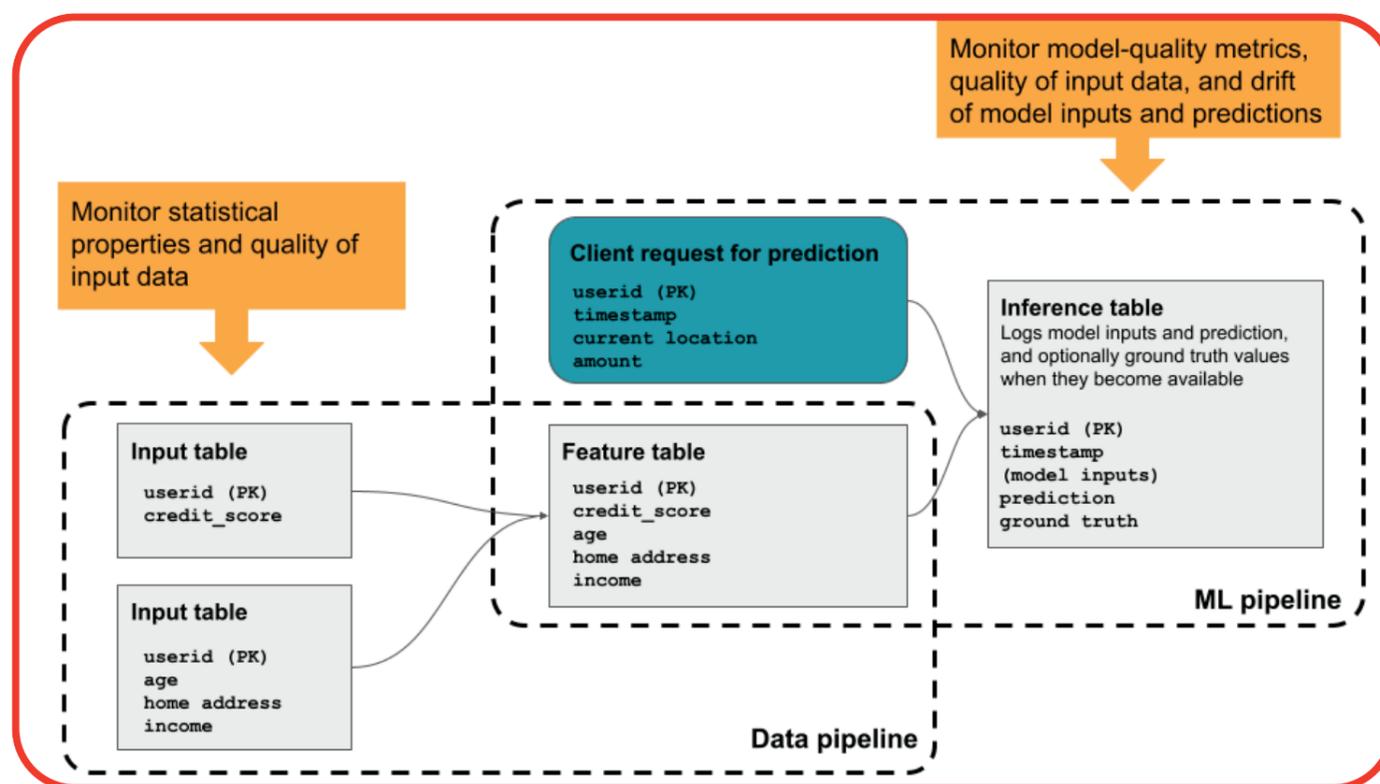
Built-in, queryable tables that provide operational metadata and insights into Databricks workspace activity.

Lakehouse Monitoring

Comprehensive tools and dashboards to observe, troubleshoot, and optimise the lakehouse architecture health and workloads.

Data Profiling

It provides summary statistics for a table, computing profiling metrics over time so you can easily view historical trends. It is helpful for in-depth monitoring of all key metrics for select tables. You can also use it to track the performance of machine learning models and model-serving endpoints by profiling inference tables.





Accelerators

Lakebridge

A unified tool that automates and simplifies the migration of legacy data warehouses and ETL workloads to Databricks SQL.

Solution Accelerators

Prebuilt, customizable templates and best practices designed to help organisations quickly solve common industry and data analytics challenges on the Databricks Platform.

External MCP Servers

Model Context Protocol (MCP) servers provide tools and APIs that AI agents can use to interact with external systems. In Databricks Marketplace, you can install external MCP servers to use with AI agent development in Databricks, including [Agent Bricks](#) and [AI Playground](#).

Private Exchanges

It allows you to make certain data products discoverable only to a specified group of consumers in Databricks Marketplace.



THANK YOU!!!



REPOST
TO YOUR
NETWORK



SHARE
YOUR
THOUGHTS



Lets
Connect!

