hiflylabs

Apps Factory

Build, run, and monitor 10, 50, or 1,000+ agent-native apps on Databricks with enterprise reliability, at startup speed.

Templates, quality gates, readiness checklists, cost control, observability setup, and more...

Your app fleet is growing. Do you have control?

You're on Databricks. Your teams are spinning Databricks Apps. And the app count is multiplying.

Databricks Apps slashed development cycles from months to weeks. Agentic engineering further shrank the timelines to days. AI, data, and software engineers ship apps at speed, at scale, completely on their own. 5 apps quickly become 10, 50, 100, 1,000+.

But with great speed gains come even greater operational challenges.

Challenge

- Every new Databricks app without shared standards adds cost, debt, and risk that compounds invisibly.
- Self-service enablement without guardrails means no ownership, no audit trail, no cost attribution.
- Scaling to hundreds of agent-native apps without an operational layer breaks governance, security, and budget.

Risk: Shadow IT

Every app starts from scratch and becomes a one-off case:

- Running on personal accounts
- Hand-pushed to prod
- Fragmented architecture
- Unbounded spend
- No data governance
- No tracing or tagging
- Interface drift
- Security debt

Accelerated pace in app development on Databricks

Ship in
Hours
A prototype, agent interface, dashboard, or human-in-the-loop app.

Ship in
Days
A chat, workflow, headless microservice, or observability app.

Ship in
Weeks
A complex, flagship enterprise app.

Apps Factory: Acceleration without risk

Operational foundation to build, run, and govern 1,000+ agent-native applications reliably, repeatably, at scale.

What is it?

Apps Factory is a 4-week review and setup program that gives your teams everything they need to self-sufficiently build secure, repeatable data and AI applications on Databricks.

30% effort per application saved

3× faster time to deploy

4 weeks to set up

What is inside?

- Tooling
- Standards
- Checklists
- Templates
- Guardrails
- Boilerplates
- Observability
- Cost monitoring

Who benefits the most?

Business and tech leaders accountable for fixing broken processes, driving operational efficiency, and making enterprise AI work—without losing control over costs, security, and integrity.

How does it remove risk?

- Shared foundation: templates, standards, patterns, and policy-as-code.
- OBO auth, RBAC, and observability to trace every app, agent, and spend.
- Logging, security, and cost governance with a 360-degree control over app fleets.

Apps Factory is the bridge that helps teams deliver business applications faster and keep ops in control.

Ungoverned fleets turn unmanageable.

Shared standards, templates, and tooling accelerate development

Excessive governance and rigid workflows kill innovation.

What does Apps Factory do?

Apps Factory resolves the compounding challenge of growing from 5 or 10 to 1,000+ apps. It creates the environment where teams across data, AI, engineering, and business departments can independently build and deploy AI applications and agentic workflows. At speed, but without sacrificing enterprise quality and standards.

Fast time to production

Standardized app skeletons, reusable templates, repeatable CI/CD pipelines, and clear framework selection guidance remove the need to start from scratch.

Security and governance by design

Standardized OBO auth, RBAC, and service principal patterns paired with Unity Catalog policies fully aligned to app access, enforced as repeatable policy-as-code.

Clear ownership and accountability

Defined RACI, production-readiness checklists, quality gates, and lifecycle management for every app.

Consistent architecture

Unified templates for analytics apps, chat and agent apps, and Genie-powered applications. Every team builds the same way, regardless of who built the first one.

Observability and cost control

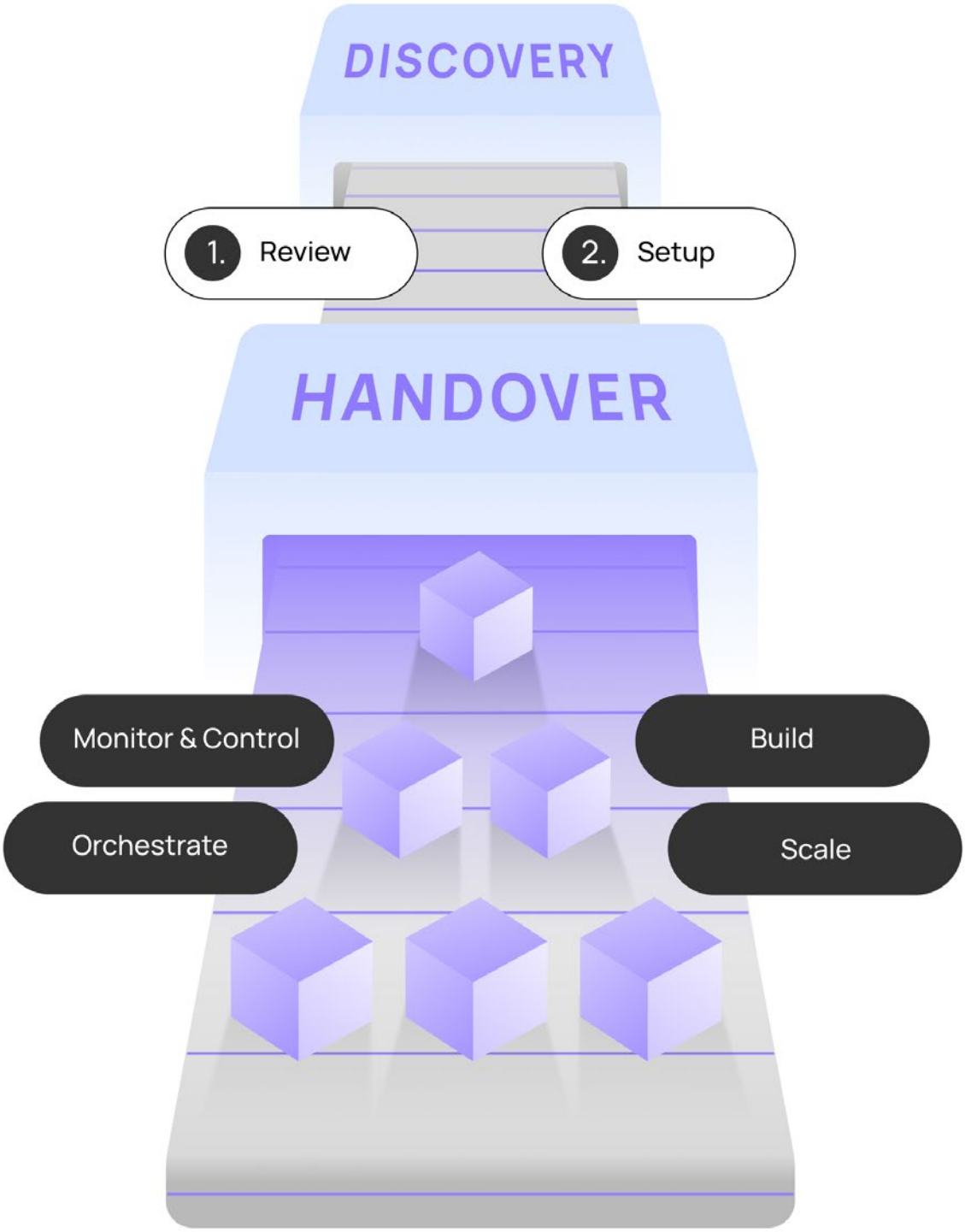
Structured logging, tracing, and usage metrics per app. Compute cost attribution is built in. Rate limits, query limits, and timeouts are enforced as guardrails so costs do not grow invisibly.

Scalable delivery model

Teams can ship many apps without multiplying governance overhead, architectural debt, or operational risk.

How does Apps Factory work?

Apps Factory runs in phases. First, we review what you have and set up the foundation your teams ship from. Then hand it to you, so you and your teams can build and run an entire ecosystem of apps—on your own and at scale.



What we build and what you own

Your operational foundation is production-ready in weeks. Let's break down how our engagement starts, what the Apps Factory setup consists of, how it's handed to your team, what it enables, and when you can start building independently.

Discovery Review Setup Handover Enablement

START – WEEK 0

We align on use cases

Discovery Workshop

HIFLY REVIEWS – WEEKS 1 – 2

We audit your environment

Architecture review

HIFLY BUILDS – WEEKS 3 – 5

The complete operational foundation

Every technical layer your ecosystem needs: standardized app skeletons, agentic engineering scaffolding, and unified governance and observability.

Your teams will never have to build from scratch again.

Architecture setup
Auto-provisioning
Templates
CI/CD setup
Security
Agentic engineering
Scaling
Testing
Unified logging
Observability
Cost control
Cost optimization
Production-readiness

HANDOVER TRANSITION

Live knowledge transfer

Handover Workshop

YOU BUILD

Your team is enabled to ship at will

AI apps
Data apps
Agentic workflows
Embedded automation

From review to enablement in 4 weeks

Review

Assessing your current setup

We run a Discovery Workshop and review to establish a precise baseline of your Databricks environment before anything is standardized or built.

- Audit across 7 dimensions: workspace topology, identity and auth, Unity Catalog governance, data and state architecture, AI and agentic components, CI/CD and observability, and cost attribution.
- Validate up to 3 business use cases and map them to the right Databricks Apps patterns.
- Identify gaps, risks, and quick wins across security, governance, and operational readiness.

Deliverables

- as-is architecture map
- maturity assessment across all 7 dimensions
- gap and risk register
- use case feasibility assessment with recommended implementation patterns
- target reference architecture
- prioritized action plan and quick wins

Setup

Building up the Apps Factory

We create the technical foundation that every subsequent app is built on: templates, standards, shared patterns, and technical governance.

- Deploy standardized app templates for analytics, chatbot, and Genie-powered applications.
- Establish shared technical patterns across all templates: secret management, structured logging, CI/CD, async jobs, MCP server, and inter-app communication.
- Hand over the full template library to your teams with an enablement workshop.

Deliverables

- production-ready app skeletons
- shared pattern library
- framework selection guidance
- local development and remote debugging setup
- OBO auth and RBAC patterns
- DABs templates
- observability and cost monitoring framework

Built once, works for everyone

Business leaders

- Deploy hundreds of agentic workflows across departments on a single AI-native application platform.
- Monitor and optimize costs for all apps and agents in one place, under centralized governance.
- Ship production-ready applications in days, not months.
- Scale without multiplying overhead.
- Keep AI within defined boundaries with built-in guardrails, preventing hallucinations and runaway agents.
- Save complex decisions and quality work in human hands, while high-volume, repetitive admin gets automated.
- Maintain high UX standards with customizable features and branded frontends.

Tech leaders

- Run a unified operational command center for every agentic workflow: observability, full traceability, auditing, and cost attribution in one place.
- Get reusable boilerplates and application blueprints so your teams never start from scratch.
- Maintain high security and governance with OBO auth, RBAC, and Unity Catalog built in, with extensions for custom workflow logic.
- Adopt a proprietary scaling framework for high-concurrency LLM calls, tested beyond default Databricks capabilities at Fortune 500 scale.
- Address source system integration, data modeling, and data quality management before anything runs on top.
- Enforce AI guardrails, overspend limits, hallucination controls, and jailbreak mitigations as policy-as-code.

Architecture

Architecture review

Architecture review is a foundational step that defines whether your current Databricks environment can scale to thousands of apps and agentic workflows without breaking or creating technical debt.

Key points to review and define:

1. Platform readiness and workspace architecture

- Databricks workspace topology
- Current usage patterns
- App deployment approach
- Runtime and supported app patterns
- Compute provisioning strategy
- Semantic layer readiness

2. Identity, authentication, and least privilege

- OBO decision points
- Dedicated service principal model per app
- Group-based RBAC strategy
- Sharing model and standard roles across apps

3. Governance and data integrity with Unity Catalog

- Catalog/domain design, ownership model, and policy scaling strategy
- Data integrity control mechanisms
- Privilege model and audit requirements
- Row/column controls and where they are required
- External data access via UC Connections

4. Operational readiness (CI/CD, observability, FinOps)

- CI/CD maturity and environment promotion
- IaC approach
- UI framework choices
- Observability baseline
- Cost attribution

5. AI, agentic and Genie components

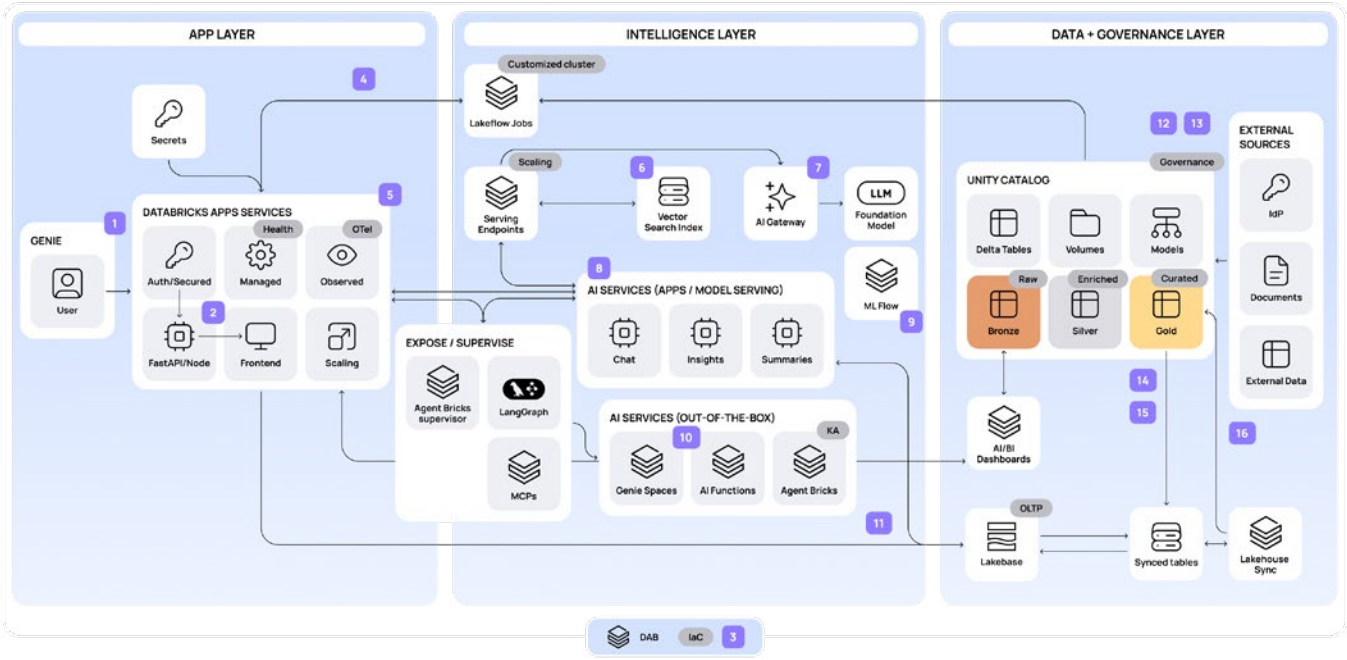
- Serving Endpoints
- Agent architecture
- RAG pipeline maturity
- Conversation history pattern
- Genie readiness
- Security guardrails

6. Data and state architecture (Lakehouse + Lakebase)

- Read access via governed tables/views
- Controlled write patterns, audit trails
- Delta table sync patterns to Lakebase
- Lakebase fit for transactional app state

Foundational architecture setup

The fewer integration points your architecture has, the fewer things can break, slow down, or create security gaps. A well-structured Databricks environment is organized into three independently scalable layers. Each is Databricks-native, each growing on its own terms without destabilizing the others.



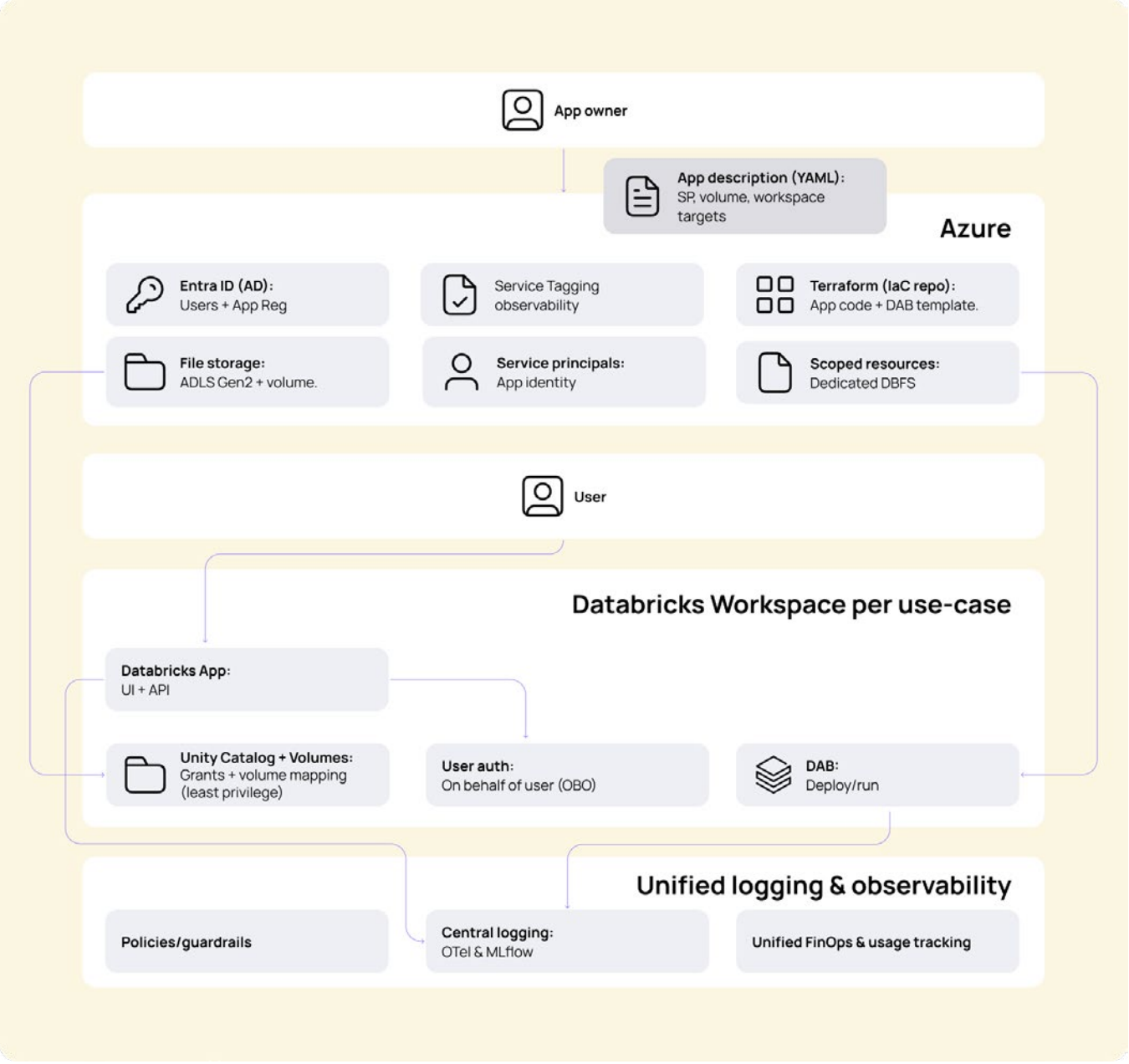
1. Business-user entry: Apps, Genie, Dashboards + Search
2. apx/AppKit
3. CI/CD templates, enforced rules. Service provisioning.
4. Unified system to online and offline data ingestion, which involves: parsing, tagging, enrichment, embedding, conversion. Utilizes AI functions.
5. Managed OLTP collector: Zerobus UC ingest - otel_ logs, otel_spans, otel_metrics
6. Hybrid search: chunks + metadata. Reranking, Evals.
7. AI Gateway centralized model access:
   - auth + rate limits + cost
   - provider routing
   - logging/governance
8. Agent runtime contract, MLflow AgentServer + ResponsesAgent
9. Inference Tables + MLflow Tracing + Evals
10. Knowledge store, benchmarks, feedback, research
11. Low-latency reads, scale to 0, branch, PITR, compute-store separation
12. Policy enforcement: ABAC (tags), RLS/CLS, audit
13. System tables:
    - Audit logs (system.access.audit)
    - Query history (system.query.history)
    - Lineage (UC lineage + lineage system tables)
    - AI Gateway usage (system.ai_gateway.usage)
    - Inference tables (payload + traces)
    - Billing
14. Keep Bronze immutable for replay. Build Silver for quality. Gold for consumption + eval sets.
15. Unity Catalog Metric Views (+ semantic metadata/ synonyms for humans and LLMs) + materialization for performance
16. Databricks/Partner Connectors, Zerobus, Lakehouse federation, Delta Sharing

Auto-provisioning

Apps auto-provisioning

Auto-provisioning removes the usual delays and bottlenecks from app creation. Any team with a validated use case can go from an idea to a deployed, governed application without waiting on IT. A single app descriptor defines everything: identity, storage, workspace targets, and deployment configuration. From there, Azure services and DABs handle the rest automatically.

App Spaces (announced at DAIS 2026) takes it further, letting admins define access and security policy once at the group level, so every app auto-provisioned into that space inherits it automatically.



Apps templates and patterns

Templates, shared patterns, and technical governance give your teams everything they need to ship without starting from scratch. Components and blueprints for every build, from a lightweight app to a complex multi-workflow enterprise system. Every new app automatically inherits the right foundation.

REPEATABLE DEVELOPMENT | REDUCED TIME TO PRODUCTION | CUSTOMIZABLE FEATURES, BRANDED UI

App templates:

Dashboard and analytics apps

- Delta table sync
- Lakebase synced tables
- Async DB connection patterns
- Row-level security
- Dashboards and visualizations

Container app with custom chat

- Router for Genie API
- Chat UI
- Inline LangGraph agent/Serving Endpoint integration
- LLM streaming
- Chat history persistence, citation, and user rating/feedback
- Conversation guardrails and source attribution

Container app with Genie integration

- Genie Space embedding/custom Genie API frontend
- Full Genie response handling
- Trusted query display, chart editing, data download
- Rating/feedback loop, send-to-review flow
- Deep research

Shared technical patterns (for all templates)

Project structure, dependency management, secret management, local development, background services, MCP server, and apps-to-apps communication.

UI framework and decision guidance

Decision matrix to choose between Streamlit/Gradio, full-stack React (with Node.js or Python/FastAPI backend), or hybrid approach depending on your use case.

Templates

Boilerplates with implementation references

Four levels depending on the complexity and how much control you need.

Level 1

Python

Streamlit / Gradio

Reference app – @Paldom

Python prototypes. Analytics, internal tools, fast delivery / demos, evaluation UIs

Skills: ai-dev-kit · databricks-agent-skills

Paldom/databricks-apps-streamlit-vibe-coding-starter

Level 2

Node.js + React

AppKit

Databricks – Product & Engineering

Managed Node.js + React SDK. UI primitives, Lakebase, Genie and other Databricks services wired up for you.

Skills: databricks-agent-skills

databricks/appkit

Level 3

Python + React

APX

Databricks – Field Engineering

Full-stack toolkit. FastAPI + React scaffold with an AI-friendly dev loop.

Skills: ai-dev-kit

databricks-solutions/apx

Level 4

Python + React Node.js

Custom full-stack

Reference app – @Paldom

Full-stack escape hatch. Your own backend, your own UI, enterprise UX, larger frontends.

Skills: ai-dev-kit · databricks-agent-skills

Paldom/databricks-apps-fastapi-starter

CI/CD

Unified CI/CD setup

CI/CD configurations and quality gates make sure every deployment in your ecosystem follows the same standardized, automated path from development to production.

Apps Factory standardizes deployment through DABs, GitHub Actions-based CI/CD pipelines, and environment promotion, with Lakebase branching built in where needed.

CI/CD pipeline and deployment with Apps Factory:

- **DABs:** bundle-based deployment as the standard
- **Environment promotion:** Dev → Stage → Prod with GitHub integration
- **Pipeline templates for all resource types**
- **Migrations:** automated DB migrations as part of CI/CD and Lakebase strategy
- **Quality gates:** automated tests running from CI/CD pipelines
- **IaC:** parameterized resource deployment

Each deployment is controlled across three independent axes. So every environment is fully isolated and promotable without manual coordination.

PIPELINE AUTOMATION | CONSISTENT DEPLOYMENTS | NO MANUAL CONFIGURATIONS

AXIS	PRIMITIVE	CONTROLS	EXAMPLE VALUES
Code	Git branch	what source ships	feature/x · dev · stg · prod
Deployment	DAB target	where & how it lands	user · dev · stg · prod
Data	Lakebase branch	which DB you read/write	dev-<user> · pr-<slug> production

Lakebase branching

Lakebase branching gives every developer or feature branch its own isolated database copy. So teams can develop and test against real data without touching shared environments.

Environment promotion approach

- Lakebase branching for isolated, per-branch database state
- Standard migration tooling for environment promotion across separate workspaces

Security

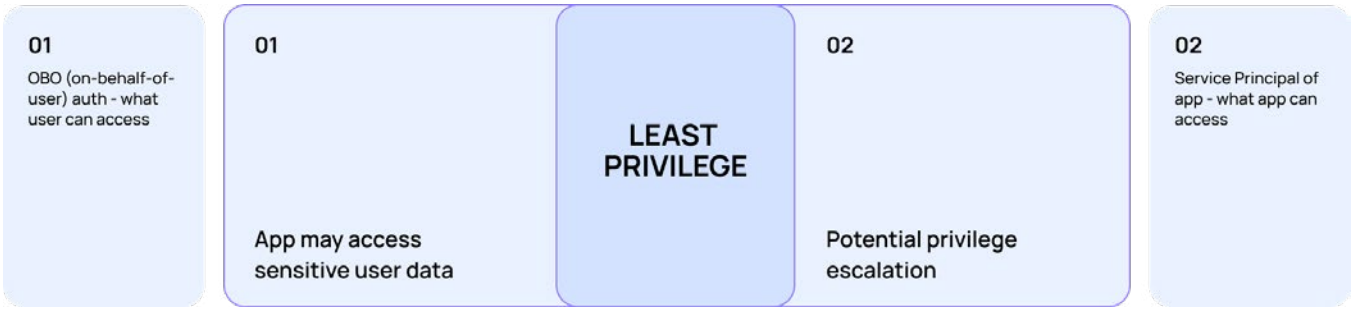
Security and access control

Standardized OBO auth, RBAC, and service principal patterns create one security and access framework applicable at every layer and enforced as repeatable policy-as-code.

Security and access control (documentation and implemented configuration):

- **OBO (On-Behalf-Of) authentication:** users interact with data through their own identity—no escalation.
- **Service Principal (SP) governance:** documented and approved SP usage per app, least-privilege principle.
- **RBAC (Role-Based Access Control):** group-based roles (CAN USE/CAN MANAGE), no direct user grants.
- **RLS (Row-Level Security):** user-based RLS in Lakebase, Unity Catalog enforced for all data access.
- **Write patterns:** validation layer → async commit job or controlled API → full audit trail, plus the ability for multi-table transaction management.
- **AI-specific security:** approved model list, tool usage policy, rate limits, prompt injection mitigation.
- **Custom access handling:** process-, workspace-, and team-based permission model that maps governance to actual organizational structures.
- **Layered automated scanning stack:** SAST (static check) + SCA (used packages) + DAST + secrets + API
- **Vulnerable & outdated components check**
- **Common web checklist:** OWASP top 10, input/output handling, etc.

RELIABLE ACCESS CONTROL | INHERITED GOVERNANCE | AI RISK MITIGATION



Guardrails and best practices for AI security and governance against: data leaks, jailbreaks, model collapse, hallucinations, and more...

Read→

Agentic engineering setup

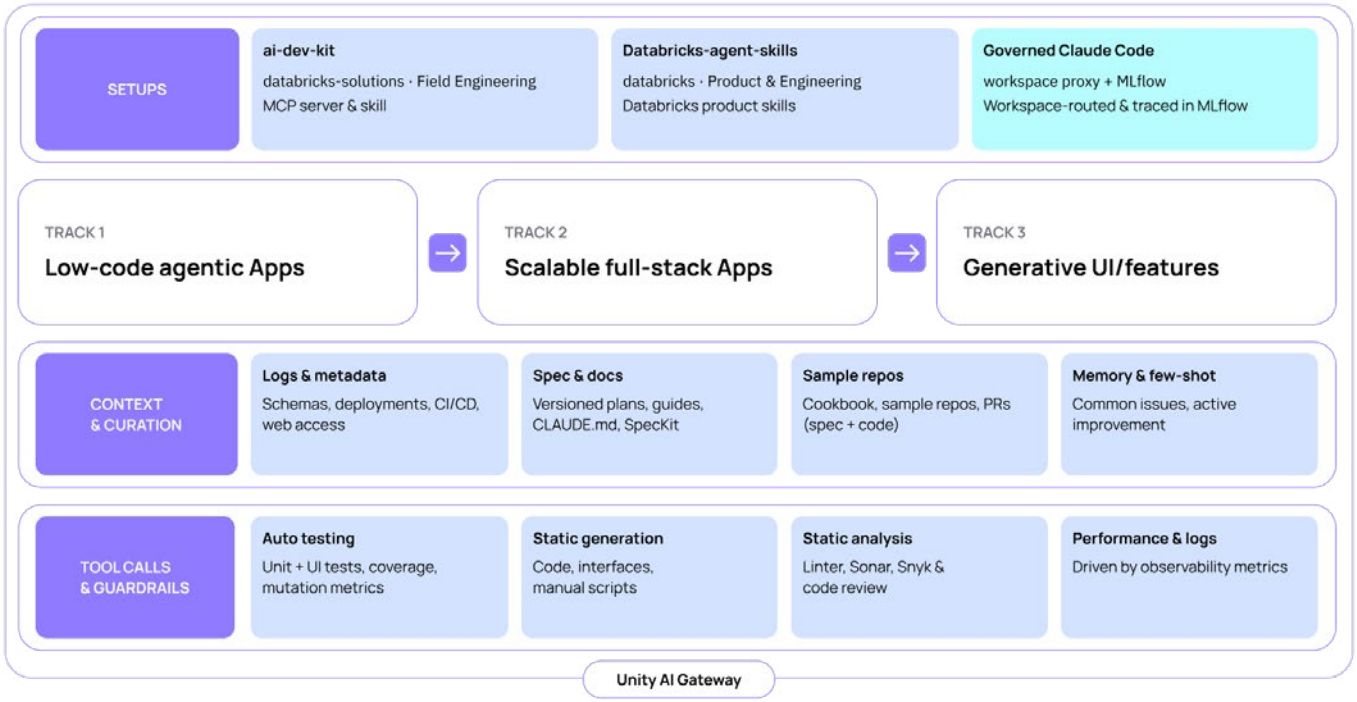
1. Databricks Apps cuts weeks of infra setup: development, provisioning, and security reviews.
2. **Apps Factory puts app delivery on a conveyor belt and makes sure it never gets out of control.**
3. Agentic engineering puts it in AI gear. You gain speed and stay on track—grounded, governed, and controlled.

Databricks Apps cut development time from months to weeks. Agentic engineering with Databricks-aware toolkits cut development time from weeks to days.

Databricks-native tools and scaffolders give your coding agents the templates, knowledge, and execution power grounded in your actual Lakehouse environment—schemas, metadata, logs, deployment states, etc. So your teams can actually trust and rely on the output.

It's governed centrally. **Unity AI Gateway** sits above every model, agent, and tool call in this setup giving you one place to enforce policy, track spend, and trace usage across your entire agentic toolkit. **Omnigent** meta-harness governs Claude Code, Codex, and custom agents in one policy-enforced layer, traced in MLflow via Unity AI Gateway.

Apps Factory includes a complete agentic engineering setup from context to guardrails with starter repos for low-code, full-stack apps, Gen UI features.



Agentic engineering tooling

Databricks Agent Skills
Official skills library for teams coding outside the workspace who need robust, guardrail-enforced Databricks guidance loaded into their assistant

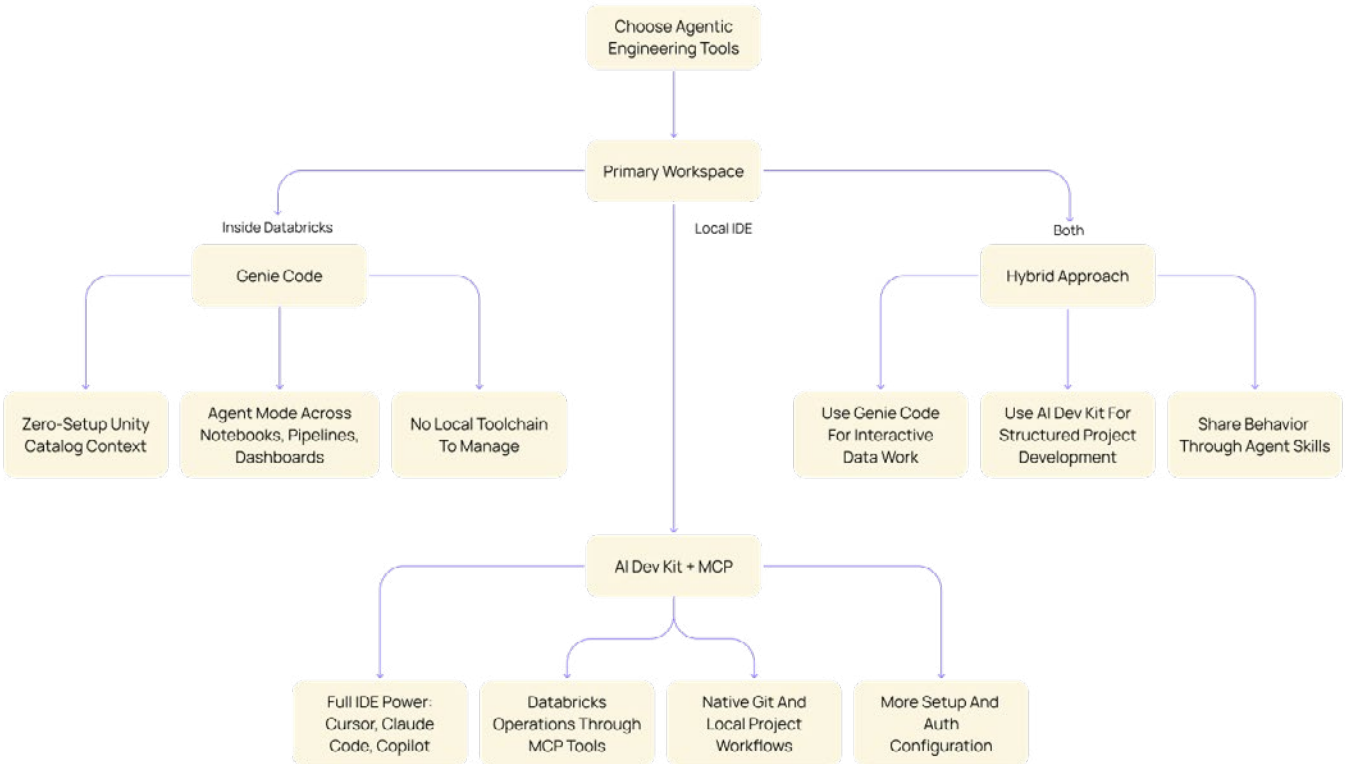
AppKit
Official Node.js + React SDK with pre-built plugins for full-stack engineers building modular, production-ready apps

Genie Code
Workspace-native AI agent for data engineers and analysts doing data-native work directly inside Databricks

AI Dev Kit
26 skills + 50 MCP tools for full-stack apps and multifaceted Databricks projects across the ecosystem

apx
Rust-based app scaffolder for rapidly generating complex, highly customized React + FastAPI applications

Agent Bricks
Official product for building production-ready agents for data teams who need governed, deployed agents fast



Scaling

Proprietary scaling framework

In most cases, apps should be architected to stay thin. Databricks provides powerful services to offload compute-heavy work to other layers: serving endpoints, jobs, and clusters.

Native horizontal scaling for Databricks Apps is now in Beta with up to 5 instances per app, built-in session affinity, zero-downtime deployments, and linear compute cost scaling. However, Beta features aren't recommended for production use, testing only.

When scaling the app layer is genuinely needed, Hiflylabs' proprietary framework enables scaling beyond current Databricks' capabilities—engineered, tested, and measured in production.

Vertical scaling

- Natively available
- Minimal instance: 70-100 QPS for a complex AI task
- **Use-cases:** Stateful apps w/ session affinity and zero-downtime deployments

Horizontal scaling

- Multiply performance by adding instances manually
- Avg. 30ms extra latency
- **Use-cases:** Long-running app functions, serving specific use cases like ingestion

How to offload work

Serving endpoints

- Unlimited endpoints, scaling automatically
- Avg. 20ms extra latency
- **Use-cases:** Large number of simultaneous (AI) operations

Lakeflow Jobs

- 10-15 second startup latency
- Customizable clusters for heavy-duty jobs
- **Use-cases**: Ingestion, document parsing, embedding, conversion, image manipulation

Testing

Unified testing

Apps Factory provides a structured, automated testing stack that covers every layer of your Databricks App, from business logic to UI behavior to production load.

Tests run automatically on every pull request. Expensive checks run before release, cheap ones run every PR.

Unified testing with Apps Factory covers:

- **Unit:** business logic, routes, and helpers.
- **Workspace unit:** fast pytest iteration directly inside Databricks.
- **Data quality:** data contracts on pipeline datasets before users see the data.
- **API smoke:** auth, app boot, and backend route validation.
- **E2E/regression:** full UI and API user path verification.
- **Performance:** sustained load, latency benchmarks, and failure thresholds.

Powered by:

- **Playwright + DAB:** deploys the app, runs UI and API regression as a Databricks job, stores artifacts in Unity Catalog.
- **Locust:** simulates concurrent users, measures p50/p95/p99 latency, generates HTML evidence reports.
- **GitHub Actions + DABs:** CI pipeline that validates, deploys, and runs tests on every PR automatically.

Layer	Tool	What it proves	App-related?
Unit	`pytest` , `vitest/jest`	Business logic, routes, helpers	☑ app code
Workspace unit	Databricks Python unit testing UI	Fast pytest iteration in workspace	workspace-adjacent
Data quality	Lakeflow, Genie code	App data is valid before users see it	data-adjacent
API smoke	Databricks App `/api/`	Auth, app boot, backend route works	☑ app runtime
E2E/regression	Playwright + DAB job	UI/API user path still works	☑ app behavior
Performance	Locust	Sustained users, latency, failures	☑ app capacity

Logging

Unified logging and governance

One Databricks-native admin view that answers across every app: **Who used it? What happened? What failed? Where was latency? What did it cost?** Achieve complete traceability from user request to agent reasoning to billing—observable and queryable from a single telemetry plane.

Every producer feeds the same telemetry plane

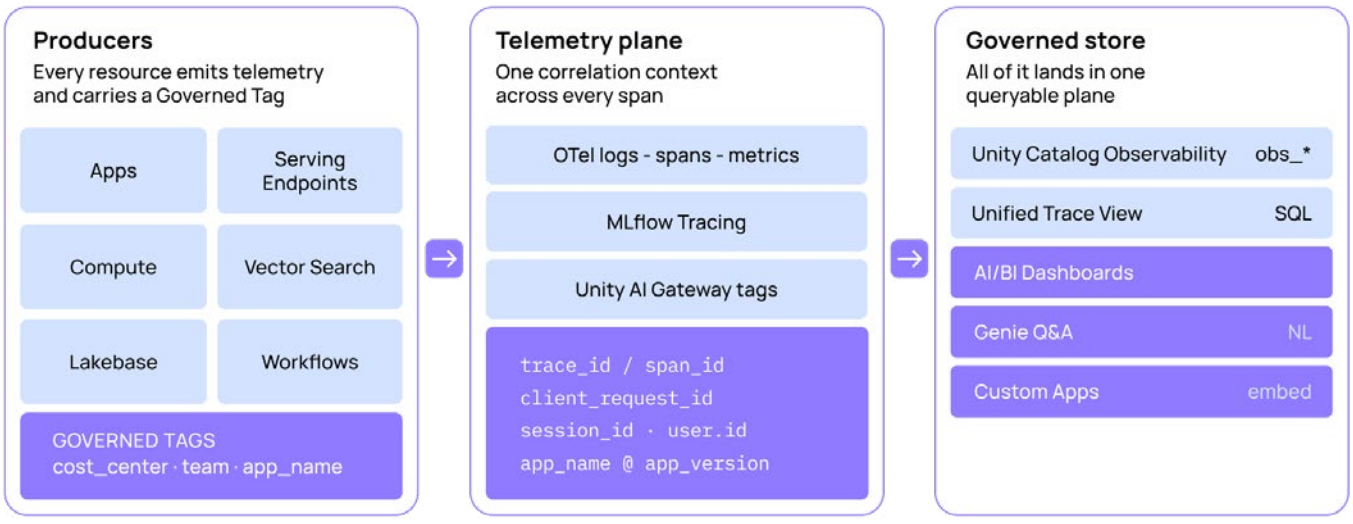
- **Databricks Apps:** OTel sidecars ship logs and metrics directly to Unity Catalog tables.
- **Custom Agents:** MLflow Tracing captures agent reasoning, tool calls, and LLM spans.
- **Model Serving + Gateway:** Unity AI Gateway tracks token usage, latency, and request tags.
- **System tables** monitor batch execution, evaluation runs, and pipeline health.

Everything flows through a unified telemetry layer

client_request_id, session_id, and trace_id go across every service for seamless joins.

All of it lands in one governed, queryable store

- **Unity Catalog Observability:** Unified tables for OTel logs, spans, and metrics joined with system audit data.
- **Unified Trace View:** MLflow trace waterfalls materialized via SQL views for rapid debugging and cost attribution.
- **Consumption Interfaces:** AI/BI Dashboards for SLOs and Genie Q&A grounded on curated silver/gold log views.



Observability

360° operational view: See every app, user, and event

A single operational command center for every agentic workflow, every app, and every cost with full traceability and auditing across the entire ecosystem.

What you can see across every app

Every component exposes unified, OpenTelemetry-compatible metrics. So health, usage, and performance are visible from one place.

Observed metrics:

- **Service health:** Heartbeat metrics and uptime status per app.
- **Performance (OTel):** p95/p99 latency, HTTP errors, and stack traces.
- **Usage stats:** DAU, session journeys, and version adoption.
- **Access control:** Audit trails for logins and permission changes.
- **Cost allocation:** AI spend split by agent, team, and cost center.
- **Agent quality:** Relevance, faithfulness, and toxicity eval scores.
- **Data health:** Freshness, null rates, and expectation failures.
- **Security:** Nightly dependency audits and vulnerability scans.

Alerting thresholds:

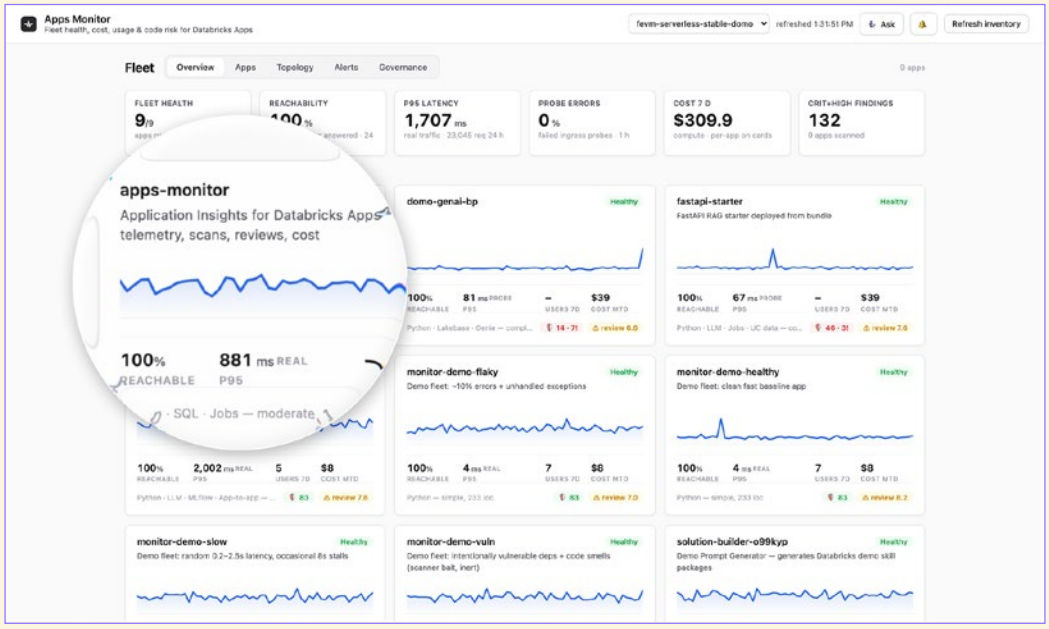
- **Availability drops:** Uptime < 99% triggers immediate on-call pages.
- **Latency spikes:** p95 shifts >2x over 7-day medians alert app owners.
- **Cost anomalies:** Daily spend exceeding 2σ triggers FinOps alerts.
- **Quality regressions:** Significant drops in model evaluation scores.
- **Data failures:** Freshness or DQ check failures in upstream pipelines.
- **Access breaches:** Suspected leakage or unauthorized permission changes.

Observability

What admins actually see

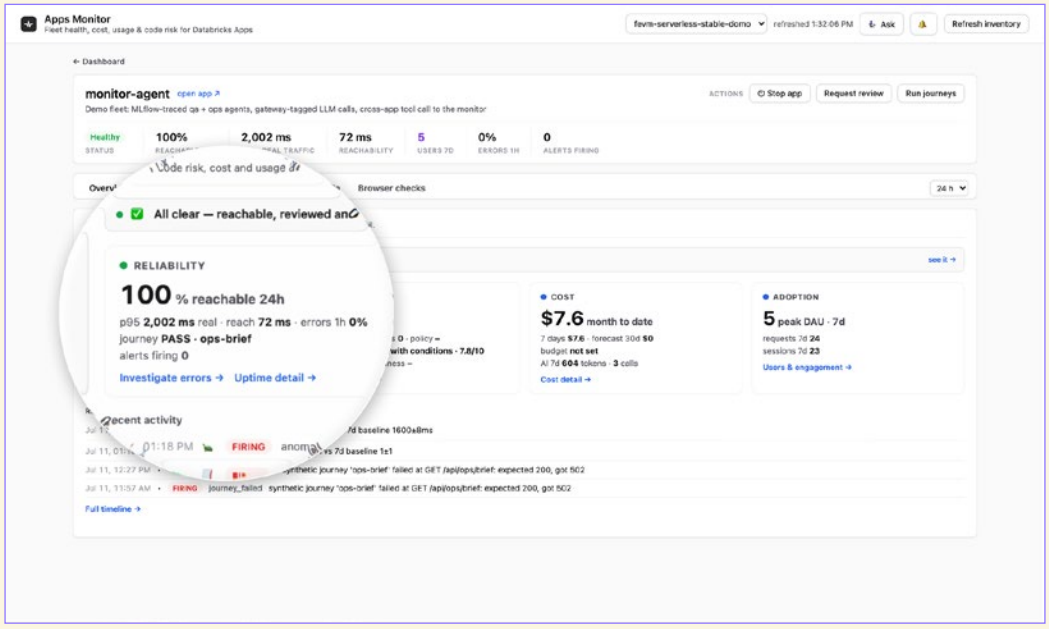
Fleet view

Reachability, real-traffic p95, users, cost and code-risk signals for every app on one screen.



Per-app drilldown

Reliability, code quality, cost and adoption on one scorecard with recent activity.



Cost control

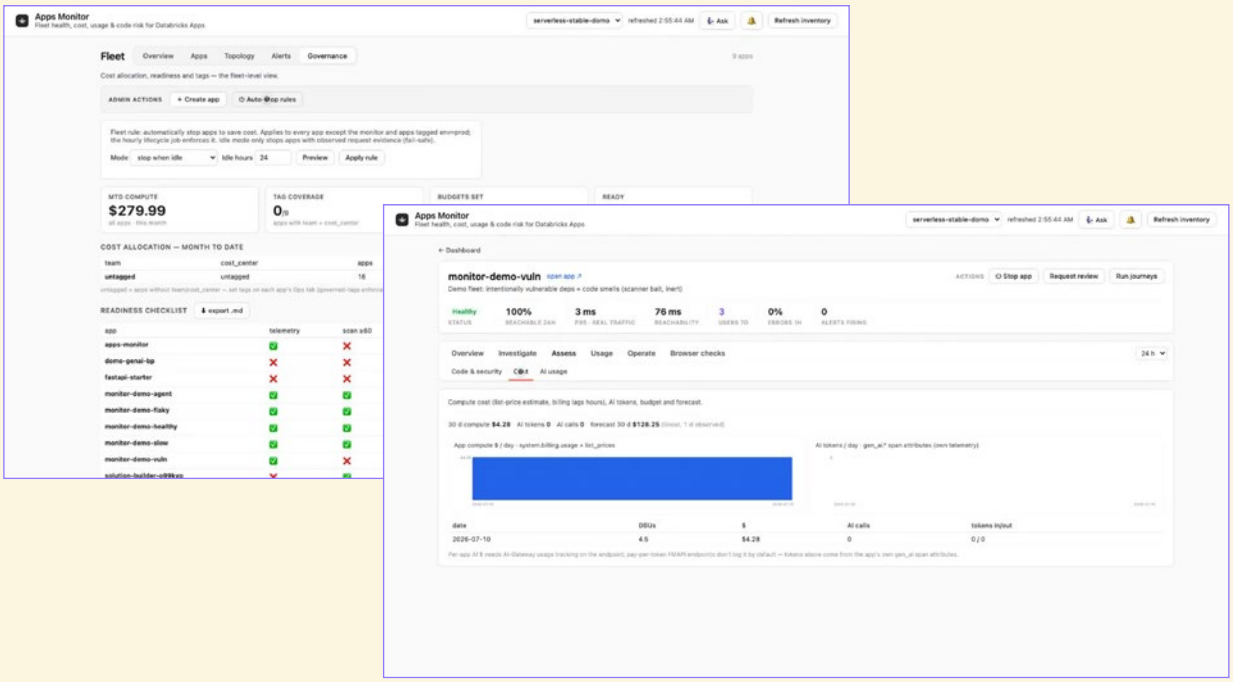
Cost monitoring

Control AI spend, trace every cent

Use governed tags to enforce cost-center visibility across every app, model call, and compute resource. So you can track all spend and prevent cost leaks before they happen

Layer	Source	Granularity
AI tokens (baseline)	your own `gen_ai.*` span attributes	per-request (exact, gateway independent)
AI tokens / latency / status	`system.ai_gateway.usage` (request_tags)	per-request (exact)
AI $	usage × `system.billing.list_prices`	per-request (exact)
App DBUs	`system.billing.usage` `billing_origin_product='APPS'`	per-app / hour
App $ per session/user	request-share allocation over OTel root spans	per-session (allocated)
Audit dollars	`system.access.audit` × billing	aggregate

Fleet cost governance



Per-app cost drilldown

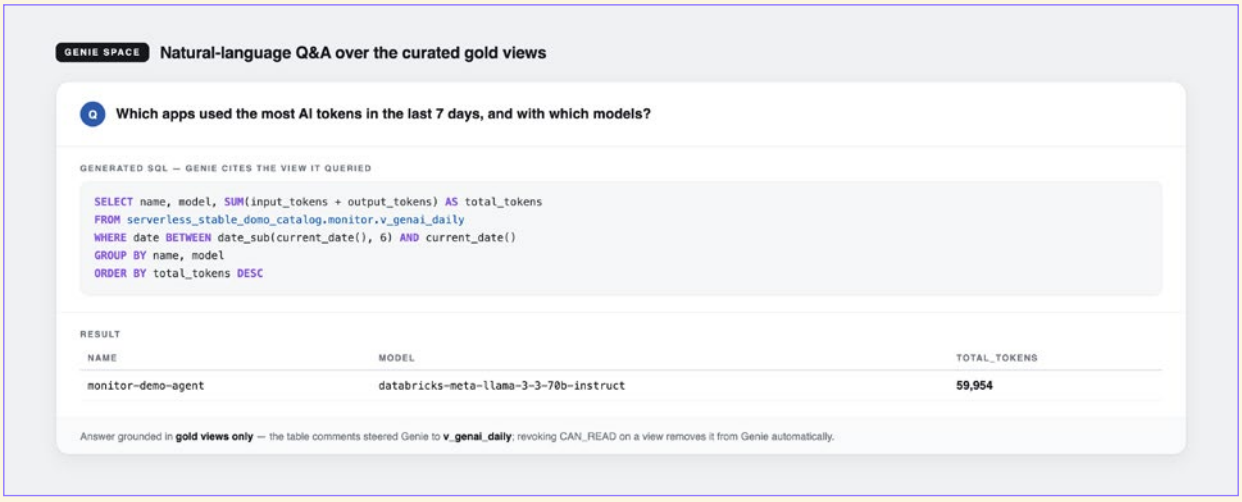
Cost optimization

Spot saving opportunities and cut spend

Apps Factory includes a custom AI/BI dashboard combined with a dedicated **cost analyst Genie Space**. Genie answers cost questions on demand by translating NL queries into SQL over the dashboard's datasets and citing the rows it used. So you can:

- Identify which teams, endpoints, or agents are driving spend and rank them by dollar impact so you know where to act first.
- Detect dormant endpoints billing with zero requests and flag them as immediate shutdown candidates.
- Compare model costs side by side and calculate projected savings from routing traffic to a more efficient model.
- Set up automated overspend guardrails to block cost overruns.
- Surface untagged workloads that make chargeback impossible and generate a prioritized fix list ranked by spend.

Instead of a number on a dashboard, you get an answer with an action plan in seconds. Ask Genie: "Which apps used the most tokens in the last 7 days, and with which models" and get the response with a dollar value in seconds.



One room for every cost conversation: all apps, agents, and compute spend in a single space.

Production-readiness checklists

One quality gate for 100% production-ready apps

The production readiness checklist is the quality gate between built and deployed. Nothing ships without passing this. It makes sure every app meets the same enterprise-grade standard, no matter who built it and what templates they used.

Production readiness + SDLC + Software architecture checklist for data and AI apps

Apps layer:

- ☑ Auth and SSO
- ☑ App access to analytical data
- ☑ Observability and monitoring
- ☑ Deployment and CI/CD environment setup
- ☑ Backup and recovery
- ☑ SLA and compliance

AI layer:

- ☑ Build (Agent Bricks functions, agents, MCPs)
- ☑ Retrieve (Vector Search)
- ☑ Serve (Model Serving + Gateway)
- ☑ Trace & evaluate (MLflow)

Data layer:

- ☑ Metadata (business/semantic description, lineage, relations, value constraints, ownership)
- ☑ Code-defined structure and migrations (incl. historical data content)
- ☑ Deployment and CI/CD environment setup
- ☑ Access control (dev / stage / prod, plus row-level security)

Non-transactional Data:

- ☑ Data Quality checks
- ☑ Load-time/Generation-time Data Quality checks
- ☑ Pipeline monitoring (operations, failures, and performance) and notification/alerting
- ☑ Auditability tracing (for ingestion and transformations), including history

Software development lifecycle:

- ☑ Global features
- ☑ Data and AI integration
- ☑ Auth and security
- ☑ Databases
- ☑ Documentation
- ☑ Source code version control
- ☑ Deployment
- ☑ UI
- ☑ SLA support and maintenance
- ☑ Scale and user
- ☑ Monitoring, testing, logging, and error handling
- ☑ Other non-functional requirements
- ☑ Traffic
- ☑ Cost and budget
- ☑ Backup and recovery

Handover

Handover workshop: own it, run it, scale it

Apps Factory equips your teams with everything they need to build, scale, and operate hundreds of agents and apps on Databricks. We equip you with everything you need for easy enablement and smooth adoption.

Handover workshop: Live walkthrough across all deliverables with adoption playbook, knowledge transfer, and Q&A

- **Reference architecture diagram:** visual overview of the entire Databricks ecosystem your organization now runs on.
- **Framework selection guide:** decision tree for when to use Streamlit/Gradio vs React/FastAPI vs full Node-based scope.
- **Production readiness checklist:** gate criteria every app must meet before promotion to production.
- **Agentic engineering guide:** setup and tools to accelerate app development inside your environment.
- **Developer onboarding guide step-by-step:** clone template → configure → deploy → monitor.
- **Supplier requirements document:** standards for external development partners covering code quality, testing, and CI/CD compliance.
- **Governance and policy-as-code:** lifecycle management rules codified and enforceable across every team and supplier.

Now you have the standards, the tooling, and the knowledge to build real agentic automation at scale—secure and governed by design.

You can spin up apps and automation workflows in days. Cut 90% of your team's menial, repeatable work. So they can focus on the remaining 10% where core decisions happen: complex judgment calls, reactive problem-solving, and client relationships.

The Apps Factory is set. Start building.

Apps range

What can you build with Apps Factory?

Apps Factory gives your teams the foundation to build standalone applications, fully automated AI workflows, or a combination of both.

Data apps

Dashboards, workflows, and self-service analytics built on top of your warehouse data. So your business teams query and act on live data directly, without an analyst in the loop.

AI apps

Chat interfaces and interactive applications that deliver AI outputs securely within the same governed Databricks ecosystem—guardrails and auth already in place.

AI agent workflows

Automated pipelines where AI agents handle multi-step processes end-to-end, with humans stepping in only where it truly matters.

AI automation embedded in apps

The best of both: intelligent automation running inside a user-facing application, keeping humans in the loop in more natural, intuitive user experiences.

Apps tiers

How far can you take it?

From a focused single use case to a company-wide ecosystem of apps and agents: Apps Factory supports any scale of ambition.

Tier 0

Prototype

PoC, prototype, or a single function to validate a concept or test a business idea.

Ship in hours

Tier 1

Single use case

One app or automation targeting a specific workflow or business need. The fastest way to get value on the ground.

Ship in days

Tier 2

Multi use case

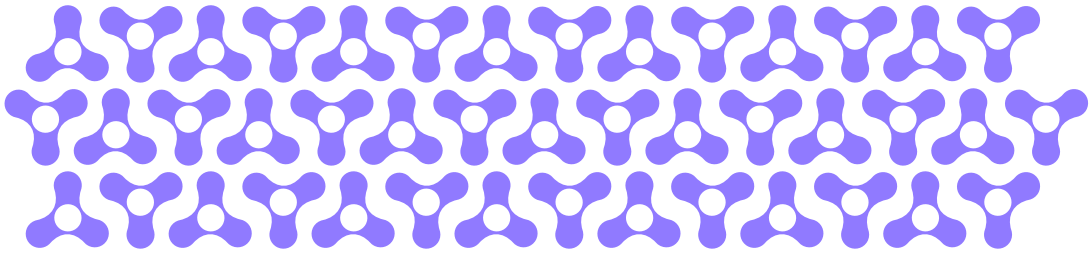
A flagship app or agentic system covering multiple related workflows, functions, or teams under one roof.

Ship in weeks

Tier 3

Ecosystem

The full realization of an AI-native enterprise: A fleet of apps and agents spanning departments and functions across the organization.



Set it up in 4 weeks

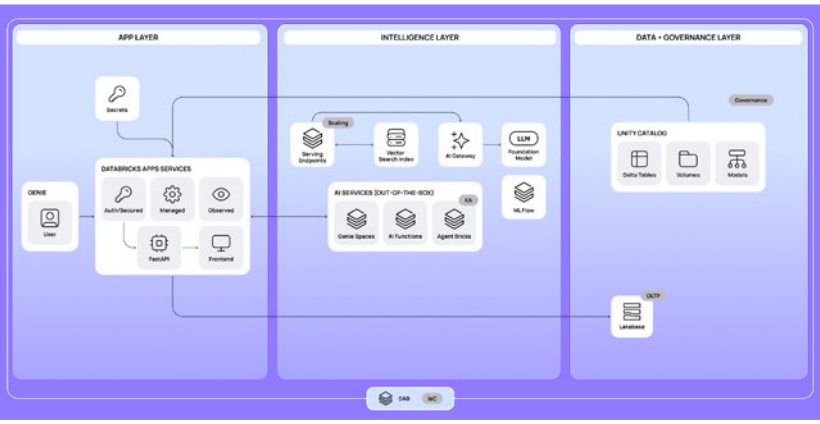
Save 30% effort per application for years

Examples

Boilerplate examples from real-world use cases

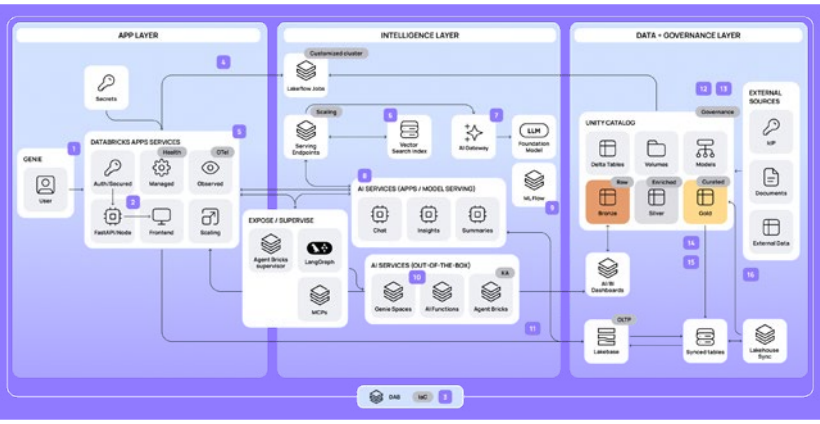
Single use case

High-performing agent-native app to automate targeted workflows and data-intensive tasks.



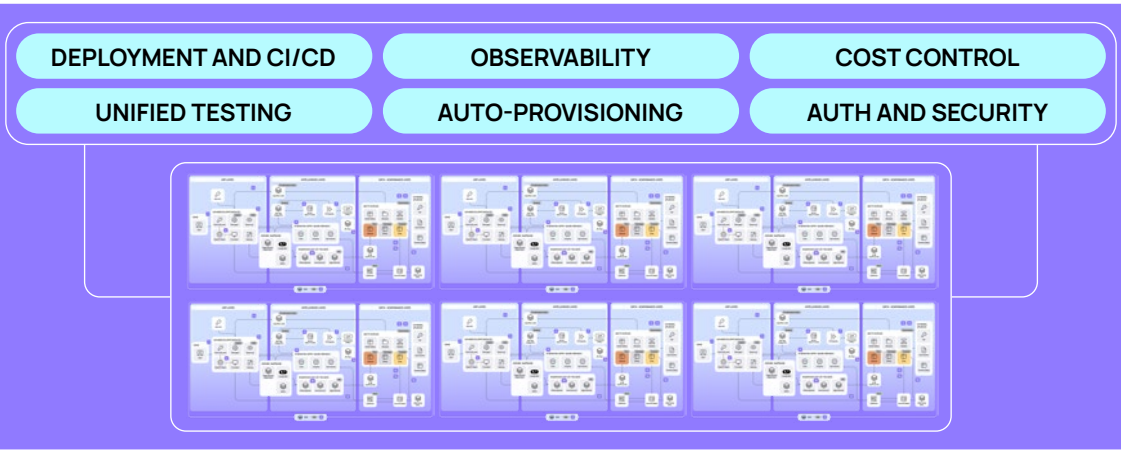
Multi use case

Flagship full-stack application with low-latency access, rich UI, and custom workflows.



Ecosystem

Enterprise ecosystem of 100+ apps with a strong data foundation and advanced AI.



Healthcare organization spins up enterprise-wide AI automation with Apps Factory

Hero

Integrated healthcare system including 7 hospitals and over 12.3M annual outpatient visits. The organization performs diverse functions across clinical care, healthcare research and education, running complex corporate operations.

Operational challenge

Despite a mature centralized digital backbone and shared IT governance, the organization lacks a standardized framework to rapidly automate high-effort manual processes and gain operational efficiency at scale.

Approach

The client teamed up with Hiflylabs to define high-potential automation use cases across workflows and departments and set up templates, tooling, and an operational layer to build and deploy applications and AI automation at scale.

Result

4 flagship use cases

40% doctor time saved on admin

Selected use cases include

- **Patient 360:** Integrated patient data and journey, surfacing decision support, medical data summaries with full traceability, and history 24/7 without manual lookup.
- **Automated cohort finder:** Processes records at 90%+ accuracy and reduces cohort build time from weeks to minutes. So research teams get to insights faster without data team bottlenecks.
- **Legal and contract risk management:** Cuts vetting time, flags risky clauses automatically, and gives full visibility into vendor and doctor contracts.
- **Supply chain management:** Predicts shortages, analyzes physician preferences, and monitors vendor contracts, eliminating last-minute procurement and supply gaps.

Set up your own Apps Factory

Your second-in-command to build and run a fleet of agent-native apps

Start with a Discovery Workshop to see how Apps Factory can accelerate development, set up fail-proof operations, and kick off safe and controlled agentic automation at scale in your organization.

Book your Discovery Workshop

No strings attached. No obligations to move on if it's not what you're looking for.

The team behind Apps Factory

Created by Databricks Apps pioneers

About the author

Domonkos Pál is Head of Digital Products at Hiflylabs, recognized for architectural thinking, fast tech adoption, and production-grade delivery standards on the Databricks platform.

- World's first Databricks Apps MVP (2025–2026)
- Member of the Databricks Product Advisory Board
- Works directly with the Databricks Apps and Lakebase product teams
- Speaker at DAIS 2025–2026 and leading data and AI events across Europe and the US

About Hiflylabs

Hiflylabs is a certified Databricks delivery partner that helps global enterprises turn data into decisions, AI into ROI, and pilots into production since 2015.

- 250+ data and AI experts in-house across offices in the US, UK, DACH, and Hungary
- 2 in-house Databricks MVPs, 50+ Databricks certifications
- Developer of the largest Databricks Apps projects running live at Fortune 500 scale
- End-to-end FDE partner: Forward Deployed Engineers with production experience

Author: Domonkos Pál
Contributors: András Zimmer, Gergely Kocsis, Zsombor Földesi, Péter Varga
Editing: Márk Lőrinczy, Alexandra Bondina
Design: Anna Füzes, Kristóf Turzó

We make AI, data, and digital products work

Domonkos Pál

Databricks Apps MVP

DOMONKOS.PAL@HIFLYLABS.COM

LINKEDIN

Hiflylabs

Databricks FDE partner

HELLO@HIFLYLABS.COM

LINKEDIN