databricks

EXECUTIVE BRIEF

How Tech Industry Leaders Are Putting AI to Work

Based on Hundreds of Databricks Deployments

Introduction

The data platforms that got you to scale are now the bottleneck for AI.

Most large tech and AI companies built their infrastructure to move fast — streaming pipelines, event systems, data warehouses stitched together over years of growth. That stack worked when analytics was the output. It breaks when AI features are the product.

Every AI capability your team wants to ship requires fresh, governed data across product telemetry, user behavior, and operational systems. When that data is siloed — or arrives hours late after ETL — your AI outputs are only as good as yesterday's batch. The result: demos that impress and pilots that stall, because the data powering them can't keep pace with the user experience they're supposed to improve.

The tech companies pulling ahead aren't building more infrastructure. They're unifying their data and AI stack on a single platform that eliminates the gap between where data is created and where AI operates on it. Eliminate the architectural tax now by unifying your data foundation, or accept a competitive gap that compounds with every quarter you wait.

Key takeaways

Multi-agent architectures, not single-purpose chatbots, are the production standard for leading companies.

Mirakl orchestrates multiple specialized AI agents to interpret product data, enforce retailer rules, and coordinate catalog workflows autonomously — reducing supplier onboarding from 28 days to under 24 hours.

Fresh data is the prerequisite for AI that works in production.

AI features degrade when they operate on stale data. Lakebase eliminates this problem by providing transactional database capabilities natively on the lakehouse — no ETL lag between your operational layer and the AI that depends on it. For tech and AI companies where differentiation depends on real-time personalization and behavioral intelligence, this is the difference between an AI feature that lands and one that doesn't.

Governance isn't a compliance checkbox — it's what allows you to scale AI across every t eam without breaking.

Unity Catalog's fine-grained access controls, column-level security, and full lineage through agent workflows let Superhuman safely scale analytics across 5 billion daily events — cutting time to insight from 4 hours to under 15 minutes.

Speed to production is measured in weeks, not quarters.

Zapier and Mirakl are examples of tech companies reaching production in weeks, not quarters. The companies moving fastest had one thing in place before they started: a governed data foundation that agents could actually use.

Enterprise-wide AI activation happens faster than most companies expect.

In under six months, every business unit at Workday launched sandbox workspaces and began delivering agentic applications. Workday teams across sales and marketing teams could now analyze product telemetry, surface performance insights, and activate data-driven campaigns in real time.

The efficiency gains are real and attributable.

Atlassian accelerated query performance to sub-minute analysis across more than 21 billion security events while reducing data ingestion costs by 80%. Superhuman cut ML data integration projects from three months to two weeks using Lakebase. Block reduced data sharing time from days to seconds across its 12PB data estate.

Open standards mean you never have to re-platform.

Native support for ChatGPT, Claude, Gemini, Llama, and every major open source model means no model lock-in — and no rebuild when a better model ships next quarter. Delta Lake, Iceberg, MLflow, and Delta Sharing are open standards. Your data assets are yours, regardless of how your stack evolves.

6x	Faster ML Delivery
80%	Reduced Costs
90%	Code Written by AI

Use cases

Multi-Agent Architectures

Your products generate data across hundreds of touchpoints — user behavior, transactions, support interactions, product events. Multi-agent architectures enable AI systems to reason across all of it simultaneously, with specialized agents handling different data types and a supervisor coordinating end-to-end workflows.

Mirakl applies this through its Nexus platform, where specialized agents reason over catalog, merchant, and retailer data — enabling automated product onboarding, intelligent discovery, transaction execution, and post-purchase support at marketplace scale.

Built on: *Unity Catalog (access controls and data lineage), Genie agents (structured data/SQL), RAG agents (unstructured documents), function-calling agents (Vector Search), Databricks Model Serving*

Natural Language Data Access and Analytics

Engineering and product teams at tech companies shouldn't be gated by analyst queues to understand what's happening in their product. Genie gives any team member — regardless of SQL fluency — the ability to query product telemetry, user behavior, or operational data in plain language.

Atlassian deployed Genie to enable security analysts to run complex queries across billions of telemetry events without writing SQL — moving from question to insight in seconds rather than hours. This approach accelerates investigations and lowers the barrier for every analyst on the team.

Built on: *Databricks Model Serving, open source LLM support (Llama, Mistral), Genie text-to-SQL, Agent Bricks AI critique agent framework*

Real-Time AI Features with Lakebase

AI features that operate on yesterday's data lose to competitors whose models see what's happening right now. Lakebase provides transactional database capabilities natively on the lakehouse, eliminating the ETL lag between your operational data and the AI layer that depends on it. For tech and AI companies building real-time personalization, behavioral recommendations, or fraud signals, this is the difference between an AI feature that works and one that ships but doesn't land. No separate operational database. No data sync. No staleness.

Superhuman replaced custom sync pipelines with Lakebase, cutting ML data integration projects from three months to two weeks — a 6x acceleration. On-call disruption from maintaining those pipelines dropped from three days per shift to roughly two hours. With Lakebase handling the data layer, engineers stopped maintaining infrastructure and started shipping features, reducing idea-to-deployment time to approximately 30 minutes using Databricks Apps.

Zillow built AI Mode — a conversational home-search experience — on Lakebase for persistent memory across sessions. "With Lakebase, we can immediately access low-latency data that's already governed through Unity Catalog, eliminating substantial incremental work," said Jaycen Grant, Director of Engineering, Data Platform at Zillow. The result: a 65% reduction in operational overhead for the data platform team, and the ability to stand up executive-facing analytics in under an hour.

Built on: *Lakebase, Delta Lake, Databricks streaming pipelines, real-time ML model serving*

Governance as Competitive Advantage

As tech companies grow, the governance problem grows: more teams, more data domains, more agents accessing sensitive data. Unity Catalog's built-in access controls, data masking, and lineage tracking make it possible to scale AI across the organization without creating compliance risk or requiring centralized gatekeeping that slows teams down.

Superhuman uses Unity Catalog to enforce fine-grained, role-based access controls and track end-to-end data lineage across 5 billion daily events — delivering insights in under 15 minutes (down from 4 hours) while reducing data ingestion costs to 10% of its previous infrastructure.

Block centralizes governance across its 12PB data estate, enabling secure, compliant access across 70+ teams while reducing data sharing time from days to seconds.

Built on: *Unity Catalog (column-level security, dynamic data masking, data lineage tracking through agent workflows, audit logging), fine-grained access controls for multi-agent systems*

Strategic Recommendations

Build Your Foundation

Before your next agent deployment, implement Unity Catalog. Its built-in access controls, data lineage and audit logging are what allow you to expand AI beyond your first team without creating governance debt. Start by auditing high-volume telemetry workflows (e.g., product analytics requests, security investigations) and identify where teams are blocked by dashboards, tickets or stale data. Then, run a focused two-week pilot on a real workflow: natural language access to product data with Genie, or a real-time AI feature powered by Lakebase. Define what "working" looks like before the sprint starts, and put a product owner on it along with an infrastructure team.

Scale What Works

Once a workflow is proven in one business unit, treat it as a template for the next. Databricks enables reusable pipelines, shared notebooks and governed data assets— meaning the second deployment takes a fraction of the effort of the first. Use MLflow to track model performance. Use Unity Catalog to manage access as new teams come online. The goal is a library of reusable agent components and governed data assets, not a collection of isolated pilots.

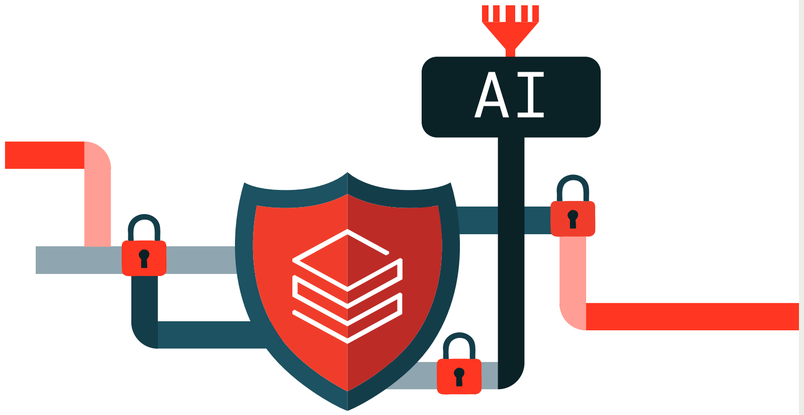
Skip What Doesn't

Not every investment is worth making. Three to avoid:

- **DIY MLOps infrastructure.** The major clouds and Databricks have already solved model serving, pipeline orchestration, and experiment tracking. Every hour your engineers spend maintaining that infrastructure is an hour not spent on serving customers.

- **Single-purpose AI point tools.** Every additional vendor increases the total cost of ownership and adds integration debt. Tech companies that win on AI speed aren't stitching together best-of-breed tools. They ship faster with the advantage of a unified platform.

- **Model lock-in.** LLMs improve every quarter. Build on an open architecture that lets you swap in better models without rebuilding your workflows. Committing to a closed model stack is the fastest way to fall behind the next capability cycle.



Results from the Field

Three patterns emerge across these deployments:

- **Speed improvements are structural, not incremental.** Block shares data in seconds. Atlassian runs sub-minute queries across 21B+ events. Superhuman cut ML project delivery cycles from three months to two weeks using Lakebase. These aren't optimizations — they're different operating tempos.

- **Efficiency compounds at scale.** Superhuman reduced on-call disruption 20x after migrating custom sync pipelines to Lakebase — from three days per shift to roughly two hours. The platform absorbs volume that would otherwise require dedicated headcount to maintain.

- **Time to production is weeks, not quarters.** Zapier (90% of initial coding and data querying handled by AI) and Mirakl (28 days to under 24 hours for supplier onboarding) show what's possible when the data foundation is already governed and accessible. Workday scaled agentic applications to every business unit within six months of platform adoption.

In every case, the common factor is the same: a platform running on Databricks that connects raw operational data to the AI layer without requiring teams to stitch together point solutions or manage ETL pipelines between them.

SUPERHUMAN	6X	Faster ML delivery Project cycles cut from 3 months to 2 weeks
zapier	90%	Code written by AI Self-serve analytics across every team
Zillow	<1hr	Executive analytics Dashboards ready in under an hour
ATLASSIAN	<1mt	Security query Speed Across 21+ billion events

Realistic timeline

When it comes to AI, business decision-makers often wonder: How long before this technology delivers something real?

Based on Databricks deployments across tech companies:

6 MONTHS

Enterprise-Wide Activation

Workday enabled every line of business to launch sandbox workspaces to build and deploy agentic applications, allowing sales and marketing to analyze product telemetry and performance and activate data-driven campaigns in real time

<1 DAY

Rapid Operational Impact

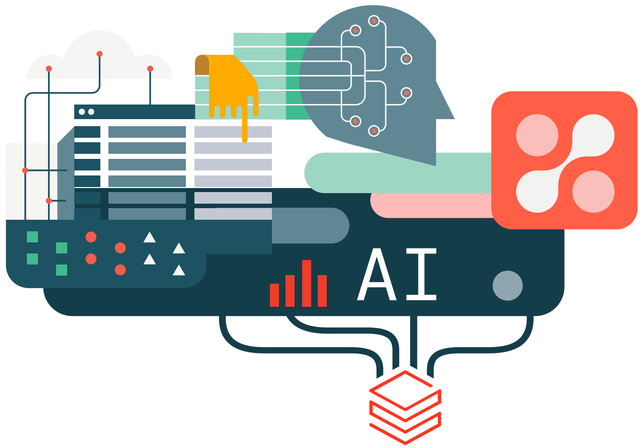
Mirakl reduced supplier catalog onboarding time to under 24 hours, down from 28 days

1 WEEK

Product-Ready Reporting

Zillow stood up executive-facing analytics in about an hour, with no engineering time

Real, measurable value can be delivered in weeks. The companies that move fastest have two things in place before they start: a governed data foundation so agents have clean, fresh, accessible data to work with, and a clear owner who defines what "working" looks like. Both are achievable before your team's first sprint begins.



Conclusion

AI is no longer a differentiator by presence — it's a differentiator by quality and speed. Tech companies pulling ahead aren't running more AI experiments; they're shipping AI features that deliver business value because they're powered by data that's fresh, governed, and accessible across every team that needs it.

Databricks is the defensible choice for tech companies at scale, and it comes down to three advantages:

Open architecture:
Native support for ChatGPT, Claude, Gemini, Llama and every major open source model. No model lock-in. No rebuild when a better option ships. Delta Lake, Iceberg and MLflow are open standards — your assets move with you.

Governance at the platform layer:
Unity Catalog makes it possible to deploy AI broadly across product, engineering, and business teams — without creating compliance risk, slowing development, or requiring centralized data gatekeeping.

A unified platform that compounds:
Every new use case built on Databricks shares the same data assets, pipelines and governance framework. The tenth deployment takes a fraction of the effort the first one did. And with Lakebase, your AI features run on data that's current — not yesterday's batch.

Speed is the final differentiator: the Databricks unified platform accelerates time to production from quarters to weeks — with the governance built in to scale AI across every team that needs it.

databricks