BairesDev Areas of Expertise | Databricks

Where our lakehouse practice is deepest, and where we would recommend to leverage BairesDev

Data engineering is our most mature practice · Spark / PySpark across 13 client data estates

	WHERE WE SUPPORT	WHERE WE'LL POINT YOU ELSEWHERE
Data Engineering & Pipelines	Our deepest capability. Ingestion and transformation pipelines, lakehouse and warehouse foundations, Spark and PySpark at scale; disconnected sources unified into one pipeline with anomaly detection and ML root-cause labeling.	Nothing structural — this is where we're strongest.
Platform Optimization & DataOps	Pipeline and job performance tuning, workspace and cluster hygiene, asset bundles and CI/CD for Databricks, restructuring domains, catalogs and schemas into a business-consumable model.	Platform procurement, capacity commitments and licensing — that stays between Databricks and the customer.
Governance, Lineage & Data Quality	Catalog and schema design, data-quality gates, validation cycles, lineage and observability so a pipeline failure is traceable to its root cause rather than rediscovered each time.	Formal data-governance audit and regulatory certification go to licensed firms. We design and build; we don't attest.
ML & AI on the Lakehouse	Feature and ML pipelines, experiment and model workflow, ML classification of data-quality failures, predictive forecasting on pipeline health, AI copilots accelerating design, testing and review.	Research-grade model training from scratch. We build, integrate and operationalize AI; we don't compete with frontier labs.
Analytics & Consumption	Dashboards and self-service analytics on governed data, ad hoc analysis across previously disconnected datasets, natural-language assistants that expose the SQL they generate and respect the user's own permissions.	Certified BI suite implementation — vendor PS or a specialized SI.
Cloud Foundations Underneath	AWS-first and deep; Azure Databricks covered. Terraform, Kubernetes, networking and identity beneath the workspace, so the platform team isn't blocked on infrastructure.	GCP-first environments — we're AWS-first by design.
How We Engage	Defined Databricks workstreams with scoped deliverables and timelines, or senior data engineers embedded inside a customer's existing platform team where they retain ownership of catalogs and pipelines.	Headcount backfill with no defined Databricks workstream. It isn't a co-sell fit and we'll tell you early rather than register it.

BairesDev

Agentic AI Success Story | Databricks

Saved 15–30 Analyst Hours a Week With an AI-Built Investigation Data Platform

Autonomous vehicle company building a self-driving fleet for mobility-as-a-service

15–30 analyst hours freed per week · 4 disconnected sources unified into one pipeline · proposal to production in ~6 weeks

About the client	An autonomous vehicle company building a self-driving fleet for mobility-as-a-service. Safety and operational decisions depend on accurate, timely data about riders, fleet behavior and safety events, pulled from many separate systems.
The problem	Every investigation started with hours of manual prep. Traffic analysts gathered rider feedback by hand, matched survey responses to operational mission data, flagged safety events and stitched together information from several disconnected systems. The work was repetitive, error-prone and delayed operational decisions.
Our solution	**Freed an estimated 15 to 30 analyst hours per week** by automating data prep and delivering investigation-ready datasets from a single pipeline that unifies 4 previously disconnected sources. **Made the automated outputs trustworthy** through repeated validation cycles, anomaly detection for volume drops, ingestion delays and duplicates, and ML classification that labels each failure by root cause. **Added predictive forecasting to flag pipelines at risk of missing their targets** and recommend fixes. **Proposal to production in about 6 weeks**, with an AI copilot accelerating the work across design, development, testing and reviews.
Technologies	**Databricks** · **PySpark** · SQL · Cursor · Qualtrics · Anthropic Claude · OpenAI
Engagement	**Staff augmentation** Client since 2024 17 senior engineers

BairesDev