Databricks Cost Optimization

"What if 70% of your workloads don't need premium compute? ***Right-size the platform. Cut costs up to 75%.****"*

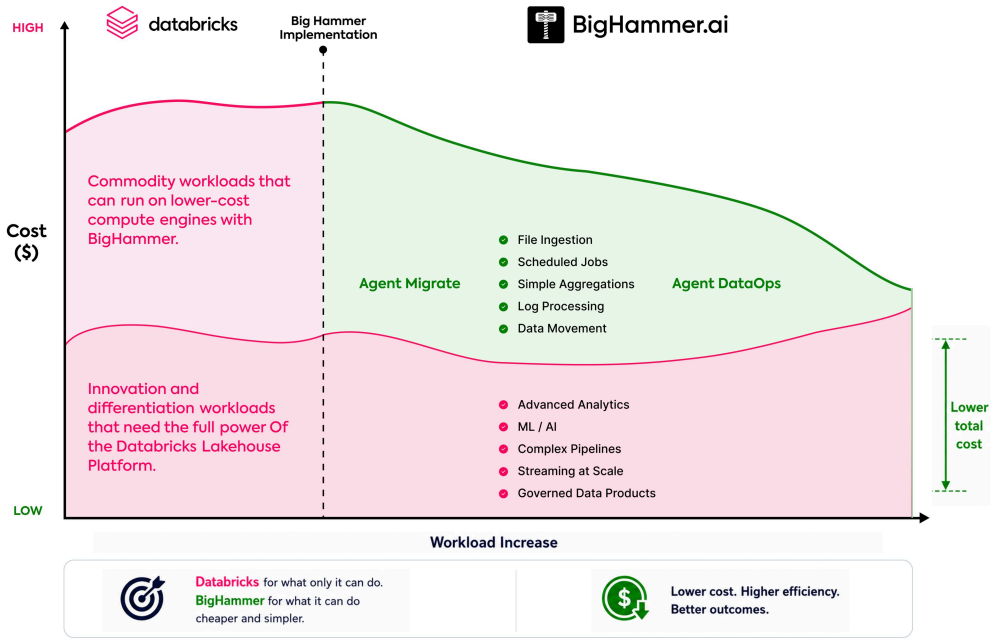
THE BUSINESS PROBLEMS

- **Shock Billing & Rising Costs-** A manageable expense is now a major budget item. Workload sprawl and unchecked DBU consumption drive unexpected increases, leaving finance teams asking: "Why are costs growing faster than the value delivered?"
- **New Business Initiatives Are Blocked-** Important Analytics, AI/ML and compliance projects are delayed or cancelled because lower complexity workloads has consumed the Databricks available budget. The platform meant to accelerate innovation has become a business bottleneck, slowing growth and delivery.
- **Operational Chaos That Keeps Recurring-** Cost spikes keep returning. Without visibility into workload costs, the same billing surprises occur every quarter. Engineering teams spend more time explaining costs than delivering value. The cycle must be eliminated, not patched.

RIGHT WORKLOAD.
RIGHT COMPUTE.

Run innovation workloads on Databricks.

Run commodity workloads on lower-cost compute.



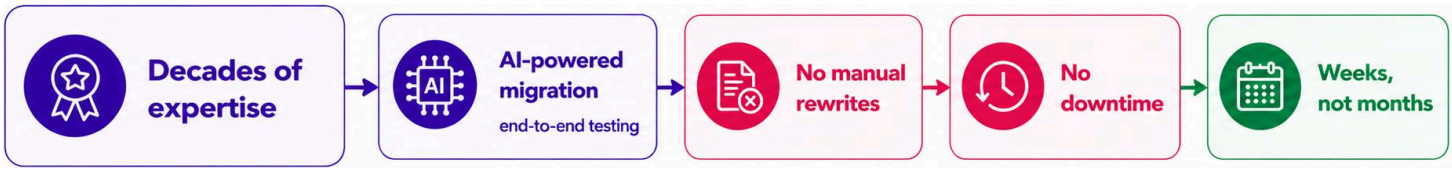
THE OUTCOME

- **TCO Focus-** Move lower-complexity workloads to a more cost-effective platform, reserving Databricks for premium analytics where its advanced capabilities deliver the greatest value.
- **Timeline & Risk Reduction-** Re-platforming is often slow and risky, trying to find cost savings before renewals and diverting engineering teams from higher-value business initiatives.
- **Change the Paradigm-** Many migrations simply move existing problems to a new platform. Inefficient code, poor monitoring and platform-dependent designs increase costs, reduce agility, and create ongoing operational overhead.
- **Use AI in the right way-** Data teams need the speed of AI, but concerns around accuracy, governance, predictability and operational control often prevent organisations from fully realising its benefits.
- **Monitor after migration-** Stop this problem before the bill arrives. Creativity monitors the workload in real time.

Databricks Cost optimization is the perfect use-case for automation with BigHammer, the data engineering platform

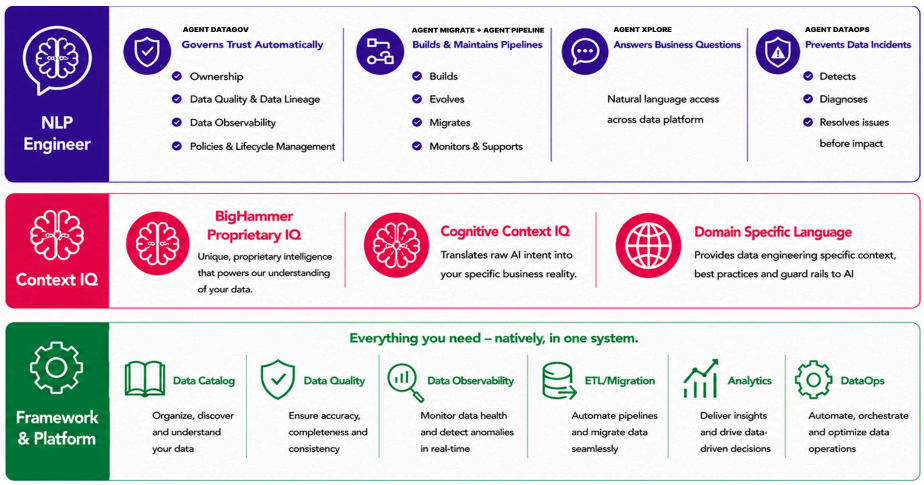
Current Scenario	Bighammer Proactive Approach
Databricks became the default platform, including simple ETL and batch jobs.	Bighammer AI agents migrates workloads to the optimal engine in weeks, not months.
40–70% of workloads don't require premium Spark clusters but still consume expensive DBUs.	Small jobs run on low-cost engines; larger workloads on Dataproc/EMR.
Delta Lake, Unity Catalog and proprietary APIs increased vendor lock-in.	Portable, platform-agnostic code with no vendor lock-in.
New workloads continue to be deployed there because it's already established.	Databricks reserved for advanced analytics and ML workloads.
Rising platform costs reduce budgets for analytics, AI/ML, and compliance initiatives.	40–70% cost savings, freeing budget for growth and innovation.

HOW IT WORKS



AI DATA ENGINEERING

End-to-End Data Engineering and Automated Data Platform Operations



KEY CAPABILITIES

There are 2 stages, this how the Agents and our platform work to support you as a data engineer to migrate workload:

AGENT MIGRATE - Migration Orchestration

Manages TCO analysis, overall workflow, and migration automation. Deploys as a container hosted in your environment, your data never leaves your perimeter. No manual rewrites, no downtime.

Agent Migrate Automation

Discover: Catalogue all workloads & dependencies

Analyse: TCO analysis, classify & prioritise

Translate: AI converts to platform-agnostic code

Validate: Parallel run with row-by-row testing

Deploy: Production deployment + CI/CD

Monitor: Continuous cost & performance tuning

AGENT DATAOPS - Post-Migration Operations

Manages and optimises ongoing costs and performance of workloads in their new destination. Full automation of support, maintenance, monitoring and cost governance for all migrated jobs.

Uses Intelligent Workload Routing

After migration: AI agents automatically route every job to the most cost-effective engine based on data size, SLA, complexity, and runtime behavior.

Example: A 2 GB daily ETL runs on DuckDB, medium workloads on Kubernetes Spark, and heavy analytics on Dataproc or EMR.

Premium Databricks: is reserved only for advanced ML workloads - executed autonomously with no manual configuration and no wasted spend.

See How BigHammer Can Reduce Delivery Time & Operational Costs

We can run an assessment of your current Databricks workloads and infrastructure to pinpoint potential cost savings, with a clear roadmap for optimisation.