

White paper

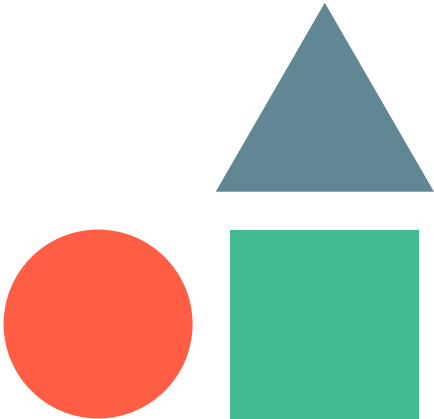
The new approach to data modernization

From migration to transformation



Contents

Introduction	3
The AI tooling challenge	4
The data challenge facing CIOs and data practitioners	5
Redefining Migration	6
How do Databricks and Google enable this and more	7
Case study	8
Conclusion	8

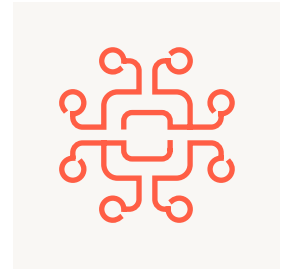


Introduction

The AI revolution is rewriting the rules on what customers expect from applications. But it is also redrawing the data workflows and infrastructure that deliver those applications.

The biggest differentiator for any organization in this new world is its data, and established enterprises should have more of this than their upstart rivals. This can give them an edge, if they can successfully access, manage and exploit that data in its entirety.

This means business leaders, technology leaders and data professionals alike need to be sure their current data infrastructure is up to the task.



The AI tooling challenge

Keeping pace with technological change is never straightforward, and today's AI landscape presents new challenges for business and technology leaders.

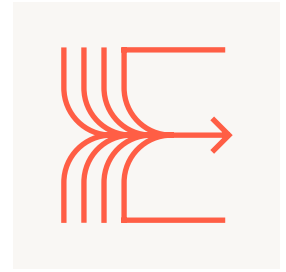
While data is the fuel, the AI tools and LLMs used to leverage it require specific characteristics and capabilities to provide concrete return on investment (ROI) to businesses.

First, successful AI initiatives aren't just about technology adoption. They're about solving specific, high-value business problems and delivering ROI. AI tools and LLMs must deliver quantifiable business benefits like cost reduction, efficiency gains, or enhanced customer satisfaction.

Second, the reliability of AI output is paramount. Tools and LLMs must incorporate robust guardrails to ensure outputs are accurate and trustworthy and explain how they arrive at decisions. These tools must go beyond mere prediction and provide genuinely actionable insights and contextual understanding that empowers better decision-making across specific domains.

Finally, AI solutions must offer seamless interoperability with existing enterprise systems and workflows. They should run on cloud infrastructure that's AI-ready and scalable, reliable, secure, and cost effective. AI tools and LLMs should free teams up for higher-value, creative work that puts an organization's ever-growing volumes of data to good use.





The data challenge facing CIOs and data practitioners

Getting the right data into those solutions is essential. But data is often held in disparate stores and geographies. It may be in the cloud, or on-prem.

Much of this data is held in legacy systems and formats, from the “big data era” or even earlier. Some will have a specific focus, such as business intelligence. And these systems will have been largely designed for structured data. They may not be able to meet the demands of today’s AI workflows and systems.

CIOs and data specialists want to deliver rapid innovation and demonstrate return on investment, which often means breaking those existing silos open, incorporating new unstructured data types, and handling much faster streams of data.

Meanwhile senior leaders will be aware of the compliance and security risks of moving too quickly and not just breaking things but losing control of their most critical data.

Data sovereignty concerns are becoming more pressing, meaning CIOs and data specialists must think carefully about where their data is held, and how it is governed.

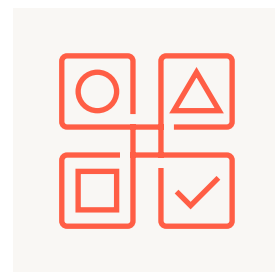
Likewise, data specialists face integration issues, disruption to their existing workflows, and the challenge of unfamiliar technologies. And often, there simply are not enough data scientists available.



At the same time, business users are increasingly asking for access to enterprise data and the tools to exploit it, further complicating governance.

How can business and tech leaders as well as data specialists meet these challenges, and ensure that their legacy data, applications, and infrastructure are reengineered so they can keep up?

And how can they justify the resources to do this, and show clear, tangible ROI, amidst relentless competition for budget, and the inherent challenge of quantifying benefits?



Redefining Migration

Data engineers have often defined their organization's data infrastructure by default, with new tools adopted from the bottom up to target specific problems.

But as success with AI becomes an existential enterprise challenge, CIOs need a more holistic approach that allows them to migrate from legacy structures to a modern data architecture, which can deliver AI powered applications and exploit the benefits of the cloud where appropriate.

AI agents can help but also raise their own challenges. Enterprises expect to deploy agents to carry out tens, hundreds, or even thousands of individual tasks. But beyond building the agent itself, this will mean preparing the appropriate data, deploying the agent, evaluating its performance, and governing the agent and its data use.

This is about more than simply swapping or upgrading individual tools, re-platforming current applications or reducing current TCO.

It calls for a holistic migration approach, spanning data and AI tooling, and the underlying infrastructure, which ultimately delivers a true AI-ready data platform, so that:



Data teams can integrate data from disparate, often incompatible, legacy systems as well as new data sources.



Engineers and data professionals have the integrated tooling they need to build unified workflows that work for them and their existing and future applications.



Data and business professionals can easily discover, manage, and access the data they need to build and feed these modern applications to deliver real business outcomes.



This is achieved with minimal disruption to existing workflows and applications, and without creating operational downtime.



Both performance and cost are predictable during and after migration.



This transformation takes account of existing skills gaps and helps both technical and business teams adapt to new technologies and tools.



Data is stored in an open format. This is key to giving enterprises full ownership of their data, retaining optionality in the future, and eliminating vendor lock-in.



Before, during, and after migration, CIOs and senior managers retain a grip on compliance, governance and security and are able to exploit the scalability and cost benefits of the cloud to ensure resilience and manage TCO.

The ultimate aim is to ensure both frontline data professionals and senior management across the organization can keep up with the dizzying pace of change AI is fueling and deliver the experiences and services their customers need.



How do Databricks and Google enable this and more

Databricks and Google Cloud are long-standing partners, enabling customers to use and integrate both companies' cutting-edge AI and data tools on Google Cloud's global, scalable, secure platform.

Databricks has transformed how organizations manage and use their data to enable both traditional applications and AI. It created the concept of the lakehouse – enabling engineers to build workflows that combine the best attributes of data warehouses and data lakes.

This approach separated storage and compute, and is predicated on open data formats, such as Apache Iceberg and Delta Lake.

More recently, Databricks has incorporated PostgreSQL into its capabilities with the acquisition of Neon. This will allow Databricks to embed a true database engine into the lakehouse architecture.

The resulting lakebase architecture will accelerate the convergence of operational and analytical data and tooling, which will be essential for the delivery of agentic AI.

This is part of Databricks' broader Data Intelligence Platform approach. Other key elements include its Unity Catalog, which gives developers and business users access to the data they need for their projects, while ensuring governance and security.

Databricks' Data Intelligence Platform uses natural language and AI to help data engineers, developers and business understand their organization's data and accelerate project and application delivery.

Databricks' strong partnership with Google Cloud gives customers access to an unparalleled array of AI tooling, including its Gemini family of models, the Vertex AI development platform, and tools such as BigQuery and AgentSpace, all on a resilient, secure, scalable cloud platform. Gemini, in an industry first, is natively integrated with Databricks' Data Intelligence Platform. This means enterprises can directly apply Google's models to their data to build and deploy AI agents, with minimal configuration and better latency and performance.

Google Cloud's underlying infrastructure includes its own custom Axion processors, which consume up to 60 percent less energy than other architectures. On Databricks operations, Axion delivers up to a 50 percent performance boost compared to other instances, while improving ROI for every dollar spent by 50 to 60 percent.

Both Databricks and Google Cloud have a commitment to openness that matches the needs of data teams who must adopt and scale new innovations quickly and can't afford to be tied down to particular data formats or models.

Together, they offer a data platform with an unprecedented degree of openness and flexibility, combined with rock solid governance and security, and virtually infinite scalability.

This gives organizations the ability to exploit the most appropriate AI tooling to extract the maximum value from their data and deliver on both functionality and ROI. This holds true whether the aim is traditional machine learning, multi agent models, or whatever comes next.

Case study



Digital Turbine is saving thousands of dollars a month, has cut delivery times by two thirds, and reduced the workload for the data management team by 20 per cent, after adopting Databricks on Google Cloud.

The digital advertising and monetization platform had previously depended on multiple technologies, tying together several ETL-based data warehouses. This caused confusion and imposed tremendous strain on its data team.

It implemented Apache Spark as its data foundation on Databricks on Google Cloud, building a data lake that could scale with the business.

This gave Digital Turbine the freedom of a technology agnostic approach, combining Google's AI tools such as BigQuery and Vertex AI, and Databricks' Unity Catalog, to maximize productivity and tackle new use cases.

Conclusion

AI and its data requirements are a challenge for enterprises but also represent a generational opportunity.

Databricks on Google Cloud provides a true platform approach that gives CIOs and practitioners the tooling, resilience, security, and scalability they need to manage their data and construct modern AI oriented workflows and applications.

This spans Databricks' collaborative data engineering, advanced analytics, open data lakehouse capabilities, and governance through the Unity Catalog, and Google Cloud's secure global infrastructure and AI ecosystem, including BigQuery, Gemini, Vertex AI, and robust serverless compute.

This partnership empowers data professionals and CIOs to innovate faster, manage risk with enterprise-grade security, and futureproof investments, helping teams deliver new AI-driven applications quickly, cost-effectively, and without lock-in.

Get Started with a Free 14-day Trial of Databricks on Google Cloud.

START YOUR FREE TRIAL

