

Maximize Data Value

Preparing your data for AI innovation

Contents

[03 /](#)

**From data sprawl to AI-ready:
A unified data foundation**

[04 /](#)

**Tackle data complexities and
challenges**

[07 /](#)

**Build a unified data foundation with
a lakehouse and OneLake**

[11 /](#)

**Unlock the power of your data estate
with Azure Databricks**

[13 /](#)

Start unifying your data for AI

From data sprawl to AI-ready: A unified data foundation

99.1% of companies surveyed say that investment in data and AI is a top organizational priority¹

AI is revolutionizing industries, automating processes, and uncovering insights at a pace that was previously unimaginable. Fueled by ample, diverse data, AI and agents can effectively identify patterns, make informed decisions, and take action. This new frontier of AI-powered systems of intelligence requires trusted, governed, current data for copilots, agents, and decision workflows to perform reliably.

Yet many AI initiatives stall for familiar reasons. Your data is split across domains and tools. Teams maintain multiple copies. Refresh cycles lag behind the business. Definitions vary by function. When data is siloed or stale, analytics and AI outputs become inconsistent. The new mandate is a simple, impactful way to cleanse, integrate, and enrich organizational data assets.

Meeting that mandate requires a unified data estate. Microsoft Fabric is the AI-powered solution that unifies experiences across data engineering, data integration, data science, real-time analytics, and business intelligence in an optimized software as a service (SaaS) environment. At the core is OneLake, a unified data lake for your entire organization. Unifying your data estate through OneLake helps reduce data duplication and sprawl while making insights available in the flow of work.

With unified, governed data and shared business meaning, you can power a system of intelligence. You can customize AI copilots that help teams explore, build, and act using natural language. Or build AI agents that work alongside people to monitor conditions, automate tasks, connect tools, and drive event-based workflows. Real-time insights and automation keep copilots and agents grounded in the latest trusted data, so they can recommend or initiate actions in the moment. AI-powered tools built into the platform help teams get more done in less time.

About this e-book

This e-book is for data and analytics leaders and technical decision-makers who want to modernize their data estate and prepare it for scalable analytics, AI apps, and agents. It helps you maximize the value of your data by:

- Unifying your data before scaling AI.
- Using OneLake to simplify complexity.
- Building and scaling an AI-ready platform.

Tackle data complexities and challenges

It's tempting to push AI innovation forward as fast as possible. But you'll get better outcomes when you start with the right foundation. That's because many data estates were optimized for reporting and transactions, not continuous, cross-domain AI use. As a result, you may be managing growing complexity as you build and scale models and analytics across disconnected systems. However, you can start addressing core data challenges with two steps: unifying data and reducing duplication.

92.7%

of firms report that interest in AI has led to a greater focus on data²



The need for unified data

To deliver reliable BI and AI outcomes, you need clean, accurate, consolidated data. When data is spread across systems, teams spend too much time extracting, moving, and reshaping it just to make it usable. The result is slower insights, higher cost, inconsistent definitions, and data that's out of date by the time it reaches the people and processes that need it.

Integrated environments stitch tools together and multiply copies, while a unified data estate establishes shared meaning, consistent governance, and reuse across analytics and AI workloads. A unified foundation helps you reduce unnecessary data movement and the friction that comes with it, including:

→ **Data consistency**

When definitions drift across teams and tools, your metrics stop reconciling and AI outputs become unreliable. A unified data strategy aligns business meaning and governance, so teams use consistent definitions across workloads.

→ **Security and compliance**

Every transfer adds security and compliance overhead. Reducing data movement helps you apply controls, auditing, and policies more consistently across your data estate.

→ **Latency**

Moving large data volumes introduces latency, delaying real-time applications and the actions they trigger. A unified platform reduces handoffs and staging, so analytics and AI run against the most current data with less delay.

→ **Model performance**

Data silos limit model training to partial or inconsistent data, slowing development and skewing outcomes. A unified foundation gives models broader, governed access to your organization's data, improving accuracy and consistency across use cases.

→ **Data transformations**

In fragmented environments, teams often repeat the same cleaning and transformation logic across multiple tools. A unified platform lets you apply transformations once on a shared foundation and reuse curated tables and definitions across workloads, reducing rework and improving consistency.

→ **Costs and resource use**

Data movement drives recurring costs—compute for transfers, storage for staging, and operational effort to run and monitor jobs. A unified platform helps lower these costs by minimizing unnecessary movement and simplifying day-to-day operations.

→ **Network bandwidth constraints**

Frequent transfers of large datasets strain network capacity and can introduce egress costs, especially with high-volume sensor data and rich media. A unified platform reduces cross-system movement so teams can work closer to where data already lives.

Reduce duplication

When the same data is stored in multiple places, duplicates accumulate as teams integrate information across systems. That duplication creates confusion about what's current and correct, and it slows your ability to deliver reliable analytics and AI because more time goes into reconciling records than generating insights.

A unified foundation reduces data duplication and its challenges, including:

→ **Accuracy and reliability**

With a single, trusted view of data, analytics and models produce more reliable results.

→ **Confidence and accountability**

Duplicate data creates a constant reconciliation tax. Teams spend time investigating discrepancies, tracing lineage, and defending numbers instead of moving work forward.

→ **Productivity and time to value**

With one trusted, governed foundation, teams onboard faster and iterate more quickly—spending less time validating data and more time shipping insights, models, and automation.

→ **Storage efficiency and retrievals**

Fewer duplicate copies reduce storage spend and keep queries fast by minimizing unnecessary data to scan and manage.

→ **Operational overhead**

A unified foundation simplifies ongoing data management by reducing exception handling, remediation cycles, and the coordination required to reconcile competing datasets.

“Microsoft Fabric was a game changer because of its ability to create shortcuts without physically moving data from one place to another. Before, if I had to incorporate three sources, I had to create pipelines to bring in the data. That pipeline had a cost. The data movement had a cost. With Fabric, it's two clicks and that's it.”

José Viegas

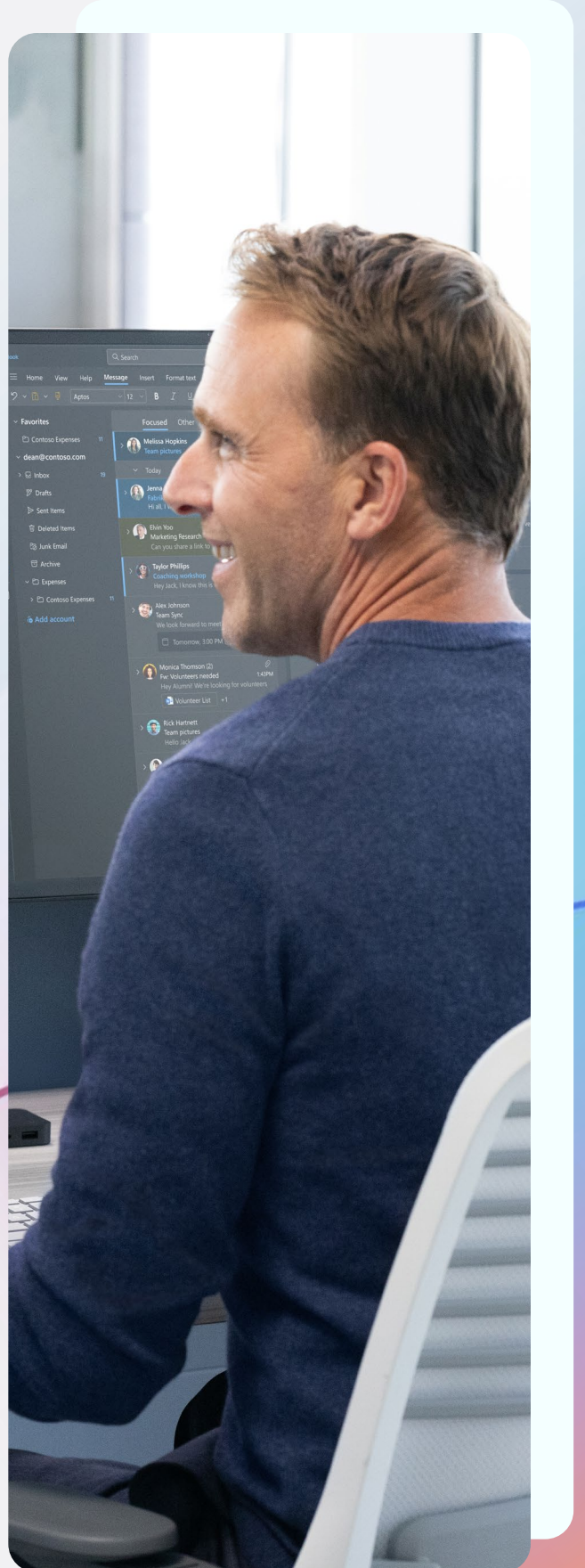
Senior Data Architect, IWG

[Read the full story](#) 

Build a unified data foundation with a lakehouse and OneLake

To reduce data movement and duplication at scale, you need a unified foundation. For example, consider an e-commerce team that adds basic personalization, then quickly outgrows it as traffic rises and customer behavior shifts. If the data behind recommendations is split across systems or refreshed too slowly, results lag and the experience suffers.

OneLake is the single, unified logical data lake for your organization. It stores data and makes it accessible across workloads. A lakehouse is the architectural pattern you implement on that foundation to bring lake flexibility together with warehouse-style management for analytics and AI.



OneLake: the foundation for lakehouse architecture

A lakehouse is a modern data architecture that lets you store structured and unstructured data together and analyze it using both Spark and SQL—without moving data between systems. Implemented on OneLake, it reduces the friction of splitting data across separate platforms, making it easier to apply consistent governance and use the same underlying data for analytics and AI.

Without data duplication or movement, you can unify your multicloud and on-premises data sources. OneLake *shortcuts* provide zero-copy access to data in external storage such as Amazon S3 and Google Cloud Storage. Fabric *mirroring* continuously replicates operational data into OneLake as open Delta tables for analytics. You can also visually connect your on-premises and cloud-based sources together with more than 150 connectors. That governed data becomes accessible across workloads, such as Data Factory, Data Engineering, Data Science, Data Warehouse, Real-Time Intelligence, and Power BI.

With OneLake as the shared foundation, you gain several advantages that help teams scale analytics and AI, including:

→ Scalable storage and processing

Lakehouses scale to support larger data volumes and more users over time. In Fabric, lakehouse data is stored in OneLake and can be processed with multiple engines, including Spark and SQL, as your needs grow.

→ Single source of truth

A lakehouse helps you standardize consistent, governed data across teams and tools. OneLake organizes lakehouse and warehouse data in a single, organization-wide data foundation.

→ Zero-copy access with shortcuts

Shortcuts and mirroring let you connect to data where it already lives and make it available in your lakehouse without creating another copy. You can reduce duplication and staging delays while keeping access governed through the platform's existing permissions and credentials.

→ Optimized for analytics and AI

The lakehouse architecture standardizes data in open Delta tables that multiple engines can use from the same source, improving performance and consistency for analytics and AI workloads.

→ Low query latency

Teams can retrieve insights quickly for interactive reporting and AI workloads. OneLake reduces latency from duplicate copies and handoffs, enabling fast access paths such as Direct Lake over Delta tables in the lake.

→ Data freshness

Lakehouses keep data fresh through streaming and frequent updates from operational systems. OneLake makes that data available across teams and workloads, through mirroring for near real-time replication and shortcuts to access data in place.

→ Data security and governance

OneLake centralizes permissions and enforces access at both the management and data levels. The OneLake catalog centralizes security management, and Microsoft Purview integration adds sensitivity labels, data loss prevention, and audit for compliance. And now you can apply role-based access down to folders or tables, including row-level and column-level restrictions.

From foundation to AI-ready capabilities

OneLake provides the shared data foundation, but preparing data for AI also requires a set of essential data capabilities to reduce movement, limit duplication, and prepare data for AI. Built on top of OneLake, these capabilities work together on the same governed data, rather than creating new silos.

Data integration

When you streamline data integration, you can reduce storage costs, simplify engineering, and improve trust in shared data. You can work with near-real-time replicas from sources such as Azure SQL, Azure Database for PostgreSQL, and Azure Cosmos DB for easy access to all of your data in OneLake.

In addition, Fabric can work with Azure Databricks Unity Catalog tables, mirroring catalog metadata and accessing the underlying data through OneLake shortcuts to avoid data movement and duplication.

Data transformation

You can transform data with low-code dataflows and automate the full process with pipelines. Dataflows in Fabric let you apply step-by-step cleaning and shaping, and pipelines schedule and coordinate end-to-end workflows. Fabric lakehouses also support Spark and SQL over the same data, so you can publish curated tables once and reuse them across workloads.

Data streaming

To act on what's happening now, you need to turn incoming signals into timely insights and automated responses. This is essential for agentic workflows, event-driven decisions, and Internet of Things scenarios, where systems detect patterns like thresholds, anomalies, or state changes and respond immediately.

Fabric Real-Time Intelligence ingests, optionally transforms, and routes real-time events without code. Fabric Activator continuously monitors streams with subsecond latency to trigger downstream actions such as notifications or Power Automate flows when conditions are met.

Data querying

Querying helps you pull the right slice of data—features, records, and time windows—so analytics and AI models are trained and validated on relevant, high-quality inputs. You can query warehouses and SQL endpoints with T-SQL, lakehouses with Spark and SQL, or real-time and log data with KQL. You can then operationalize those results through semantic models and Direct Lake for fast, governed analysis.

Data visualization and collaboration

When governed data becomes shared understanding, your teams can track outcomes, spot changes, and make decisions faster. Fabric includes Power BI as a first-class workload for interactive reporting on data in the platform. Direct Lake supports interactive analysis over large Delta tables in the lake, reducing refresh overhead compared to traditional import patterns.

Additionally, Microsoft Copilot in Fabric is available across workloads to help you interact with your data using natural language, transform and analyze it, generate insights, and create visualizations and reports across Fabric and Power BI.



“We help users unlock the value of their data by asking questions specific to their business to make better data-driven decisions.”

Srihari Kumar

Chief Product Officer, Insights and Data, JLL

[Read the full story](#) 

Azure Databases: Deliver AI apps and agents faster, with built-in capabilities for generative AI and LLM workloads

Azure offers fully managed services for SQL, NoSQL, PostgreSQL, and MySQL, providing the foundation for any size business-critical workload your organization uses to achieve AI transformation and business innovation. With Azure Databases, companies get up to five times faster performance, intelligent threat detection, and up to 58% lower total cost of ownership.³ New apps, AI agents, and data-driven solutions benefit from GitHub Copilot-assisted development, zero-touch recovery, single-digit millisecond latency, advanced vector search capabilities, and integrations with the most loved developer tools, analytics platforms, and AI technologies. With Azure Databases, your data becomes a strategic asset, helping turn your AI ambitions into real business results.

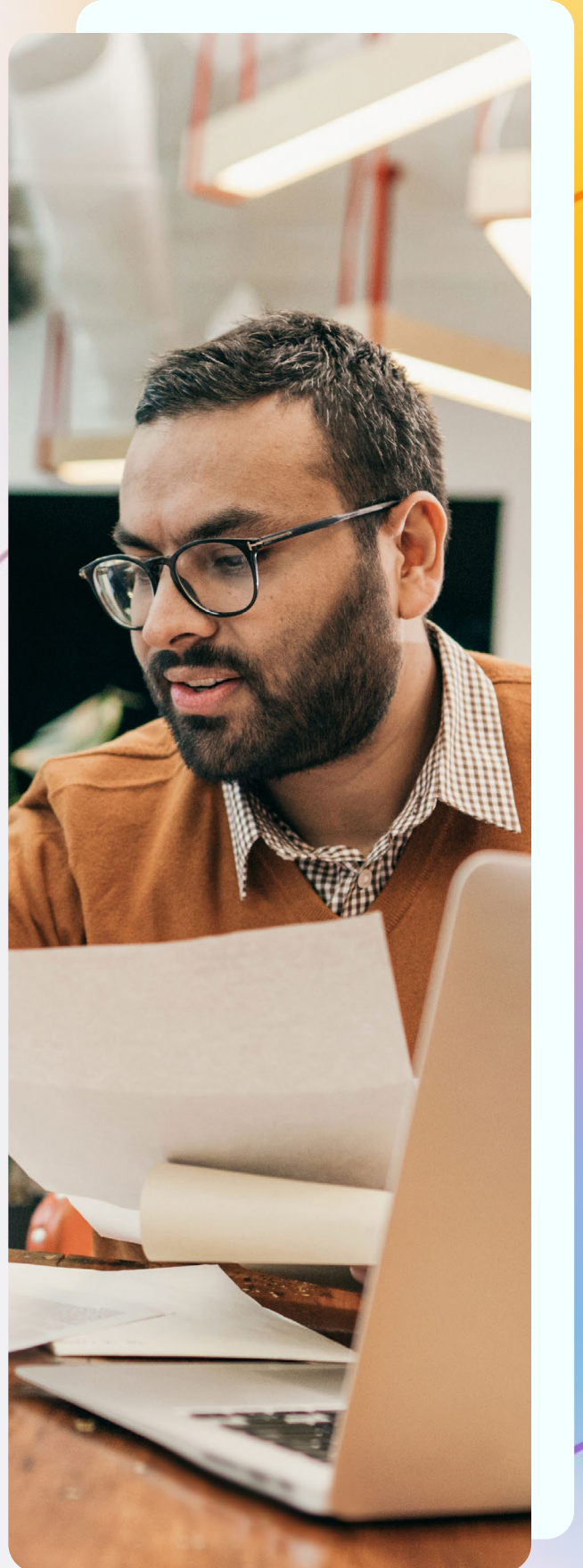
Unlock the power of your data estate with Azure Databricks

In the evolving data analytics landscape, selecting the right cloud platform is critical. Widely adopted, Databricks provides a unified platform for data, analytics, and AI built on open, trusted standards. Azure Databricks represents a unique partnership between Databricks and Microsoft. A joint engineering collaboration ensures tight integration and optimized performance, enabling you to take advantage of a truly unified data estate. Your data engineering teams can work with the Databricks Data Intelligence Platform with Unity Catalog interoperability with OneLake to accelerate open, governed analytics and AI on Azure.

Key benefits include:

→ **Seamless integration**

Azure Databricks is designed to work with Entra ID and other Azure services for identity and authentication, storage, and observability services, making it easy to set up and manage your data workloads.



→ **Scalability and performance**

Clusters can automatically scale up or down based on workload demands, for optimal resource use and cost efficiency. Azure offers a variety of compute options, including GPU-enabled instances, which accelerate machine learning and deep learning workloads.

→ **Security and compliance**

Azure holds industry-leading compliance certifications, ensuring your Azure Databricks workloads meet regulatory standards.

→ **Cost optimization**

Azure provides a unified billing experience, so you can manage Azure Databricks costs alongside other Azure services. You can take advantage of Azure Spot Instances to reduce costs for noncritical workloads, or choose Reserved Instances to save on long-term Azure Databricks usage. You also have options for monitoring and managing your costs when using Databricks AI in Azure Databricks.

→ **Integrated services**

Within the Azure ecosystem, you can connect Azure Databricks with Microsoft Foundry for advanced AI modeling, and work with Power BI for real-time visualizations and reporting, among a host of other integrations from Microsoft Fabric to Copilot Studio and Power Platform.

“We chose Azure Databricks because it works well with our existing Azure ecosystem, allowing us to leverage and extend our current investments. This meant our teams could continue using familiar systems while accessing new analytics features.”

Benedict Ho

Senior Deputy Director, DSA Department,
Central Provident Fund Board

[Read the full story](#) 

Start unifying your data for AI

For AI to deliver real business value, start with a unified data foundation that every team can trust and reuse. The fastest path to impact is reducing friction—fewer silos, fewer copies, and less time spent reconciling which version is right.

An end-to-end platform gives every team what it needs to move from raw data to AI-powered intelligence. Across roles and departments, your users can more efficiently explore data, generate insights, and build AI apps and agents, powering your organization's AI transformation with a complete data platform.

Take the next steps

- ➔ [Get started with a complete data platform using Microsoft Fabric](#)
- ➔ [Learn more about maximizing your data's value with Azure Databricks](#)
- ➔ [Kickstart AI development with Microsoft Foundry](#)

¹ Bean, Randy and Davenport, Thomas H. [Survey: How Executives Are Thinking About AI in 2026](#). Harvard Business Review. January 12, 2026.

² *Ibid.*

³ Performance and cost savings figures are drawn from Microsoft-commissioned studies and customer case studies. Actual results may vary depending on workload, configuration, and deployment context. See, for example, McAfee, Nathan. [The Economic Benefits of Migrating On-premises Databases to Azure Database for PostgreSQL](#). Enterprise Strategy Group. March 2025.