

Modernize your data estate by migrating to Azure Databricks



Contents

SECTION 1:	The data-driven transformation imperative.....	4
	Data's role in strategic growth.....	4
	Key challenges with legacy systems	4
	Modernizing with Azure Databricks.....	5
	Strategic partnership for future growth.....	5
SECTION 2:	The new data reality — Why change is necessary.....	6
	The growing need for modern data infrastructure	6
	Rising data demands require new approaches.....	6
	Limitations of traditional data warehouses.....	7
	The benefits of a modern cloud-native platform.....	8
SECTION 3:	Azure Databricks as a catalyst for transformation	9
	Streamlining data operations.....	9
	Accelerating time to insight with real-time processing	10
	Driving innovation with AI and machine learning	10
	Advantages of the Azure Databricks platform.....	11
SECTION 4:	The strategic advantage of migrating to Azure Databricks — AI and real-time analytics.....	12
	Maximizing AI and ML capabilities	12
	Real-time analytics for rapid insights	12
	Scaling data operations to drive innovation.....	13
	Enhanced value through Azure integrations.....	14
SECTION 5:	Key benefits of migrating to Azure Databricks.....	15
	Reducing operational overhead	15
	Enhancing scalability and flexibility.....	15
	Leveraging advanced analytics and AI.....	16
	Improved data governance with Unity Catalog.....	16
	Cost savings and predictable pricing.....	16

Contents

SECTION 6:	Enhancing data governance and security with Unity Catalog	17
	Centralized governance and data access control.....	17
	Data lineage for transparency and compliance.....	17
	Simplified compliance with automated governance.....	18
	Enhanced security with Azure integration.....	18
	Supporting long-term governance goals	18
SECTION 7:	Preparing for a successful migration to Azure Databricks	19
	Step 1: Assess your current data infrastructure and needs	19
	Step 2: Develop a phased migration strategy	19
	Step 3: Leverage Azure Databricks migration tools and support.....	19
	Step 4: Prepare data for migration	20
	Step 5: Test and validate migrated data.....	20
	Step 6: Facilitate change management and team training	20
SECTION 8:	Measuring success and maximizing value after migration	21
	Establishing KPIs to measure migration success	21
	Optimizing workflows and operations on Azure Databricks	22
	Driving innovation with AI and advanced analytics.....	22
	Maximizing ROI with Azure integrations.....	23
	Continuous improvement for long-term value	23
SECTION 9:	Future-proofing your data strategy with Azure Databricks	24
	Building a scalable foundation for long-term growth.....	24
	Adapting to emerging technologies and analytics trends	24
	Driving innovation with AI and machine learning	25
	Reinforcing data governance for a secure future	25
	Building security with Azure integrations	25
SECTION 10:	Driving growth with Azure Databricks	26
	Building a resilient, future-ready data environment.....	27
	Taking the next step.....	27

SECTION 1:

The data-driven transformation imperative

Data's role in strategic growth

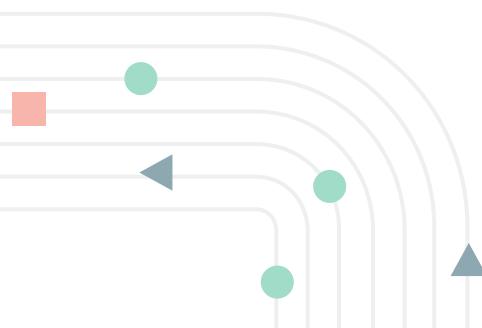
Data is integral to decision-making and operational efficiency across industries. Every day, companies generate extensive amounts of data from diverse sources, including sales transactions, social media, Internet of Things (IoT) devices, and AI systems. The convergence of traditional data sources with AI — including generative AI and large language models — has created new opportunities and challenges for organizations. The ability to process, analyze, and leverage this data effectively allows organizations to respond swiftly to market trends, optimize operations, and create new business opportunities through AI-powered insights.

However, many companies still operate with legacy, on-premises data warehouses that struggle to meet the growing demands of modern analytics and AI workloads. These systems, designed for traditional structured data processing, often fall short in delivering real-time insights, supporting advanced analytics capabilities, supporting organizational growth, and handling the computational requirements of modern AI applications, limiting an organization's ability to fully harness its data for strategic initiatives and AI-driven innovation.

Key challenges with legacy systems

Traditional on-premises data systems present several common challenges that impact scalability, cost, and performance:

- A. High maintenance costs** – Legacy systems require ongoing investments in hardware and infrastructure management. These expenses often divert resources from innovation and limit IT budgets for strategic priorities.
- B. Limited scalability** – On-premises data warehouses lack the flexibility of cloud-based platforms, making it difficult and costly to scale as data volumes grow. Fixed capacity can prevent organizations from adapting quickly to fluctuating data demands.
- C. Performance bottlenecks** – Slow data retrieval, limited support for concurrent users, and inefficient processing lead to delays and performance issues. As data volumes increase, these limitations become more restrictive, impacting decision-making.
- D. Data silos** – Integrating data from different sources and departments can be complex within legacy environments, leading to fragmented data and limited visibility across the organization. Siloed data restricts the scope and accuracy of analytics, further limiting potential insights.



Modernizing with Azure Databricks

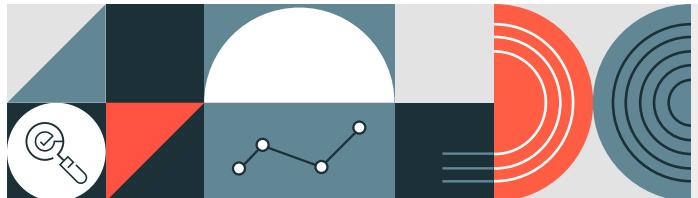
Azure Databricks offers a cloud-native, AI-powered Data Intelligence Platform co-developed with Microsoft that addresses the challenges of legacy systems. This first-party platform provides the scalability and flexibility needed to meet the demands of modern analytics, including machine learning (ML), real-time data processing, data warehousing, and generative AI applications. Migrating to Azure Databricks can deliver:

- **Seamless integration within the Azure ecosystem** – Azure Databricks allows organizations to leverage the full Azure suite, ensuring compatibility across various business functions, from data engineering to AI workflows.
- **Reduced infrastructure management** – By moving to a managed, cloud-based solution, companies can shift IT resources from maintenance to more strategic, value-added activities. This shift supports agility, operational efficiency, and cost savings.
- **Enhanced support for real-time and AI-driven applications** – The platform's capabilities in handling high-demand analytics make it possible for organizations to act on real-time data insights. With AI/Business Intelligence (BI) Genie integration, users can leverage natural language queries and generative AI features to democratize data access and accelerate insight discovery, helping drive responsive and proactive decision-making.

Strategic partnership for future growth

Microsoft and Databricks unite leading technology and innovation to support organizations in their data transformation efforts. As the first-party service within Azure, Azure Databricks is optimized for security, governance, and scalability in the cloud. Microsoft and Databricks work collaboratively to ensure that Azure Databricks evolves alongside industry needs, offering organizations a platform that is continuously updated and aligned with enterprise-grade requirements for security and compliance.

Organizations that adopt Azure Databricks can confidently transition to the cloud, knowing that the platform's features and updates are designed to support both current and future needs in analytics, data science, and machine learning.



DID YOU KNOW?

Azure Databricks provides Unity Catalog, an integrated data governance solution that allows businesses to manage data access, track usage, and enforce compliance across data environments. Unity Catalog supports transparency and security for organizations with complex data governance requirements.

SECTION 2:

The new data reality – Why change is necessary

The growing need for modern data infrastructure

Data is one of the most valuable assets for any business, and modernizing data infrastructure is critical to leverage it fully. Many companies still rely on legacy systems that limit scalability, add operational costs, and slow access to real-time insights, all of which impact competitiveness. By migrating to Azure Databricks, organizations can integrate both structured and unstructured data in real time, making it feasible to implement agile, data-driven decisions and optimize operations. Starting with a phased migration plan helps test and validate the new system, reducing transition risks and maximizing efficiency improvements.

As data demands rise, organizations are finding that traditional data warehouse systems cannot keep up. This section covers the impact of today's data needs, the challenges of outdated systems, and why a cloud-native solution like Azure Databricks is essential.

Rising data demands require new approaches

The rate of data generation continues to accelerate, fueled by a mix of digital interactions, IoT devices, and data sources that extend beyond structured formats. Businesses now need systems that can manage both structured and unstructured data — such as customer feedback from social media and real-time sensor data. Traditional systems often cannot handle this level of diversity and scale.

Key aspects of these growing demands include:

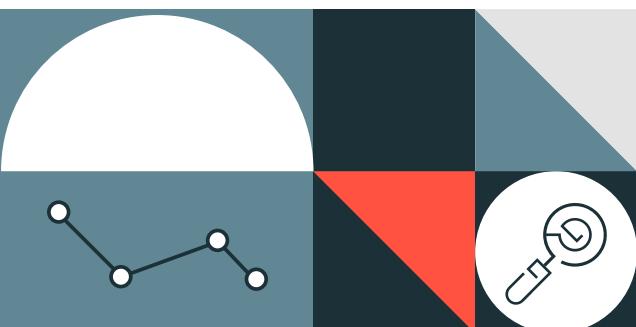
EXPANDING DATA SOURCES	REAL-TIME DECISION-MAKING	ADVANCED ANALYTICS
<p>Data now originates from numerous channels, including mobile applications, connected devices, web analytics, and transaction systems. Organizations require a unified architecture that can efficiently integrate and process these various data types.</p>	<p>Timely insights are increasingly critical for tracking customer behavior, optimizing operations, and adapting to market changes. Fast data processing enables businesses to act quickly and make informed decisions.</p>	<p>AI and machine learning are becoming standard tools for uncovering predictive insights, automating processes, and enhancing customer experiences. These complex workflows require infrastructure capable of high-volume data processing and sophisticated data transformations.</p>

Legacy infrastructure often cannot support these modern demands, creating barriers to effective data use. Azure Databricks offers the scalability, speed, and adaptability needed to fully leverage data as a strategic resource.

Limitations of traditional data warehouses

Traditional data warehouse appliances were designed for stable, structured reporting, making them ill-suited for the dynamic data landscape of today. The following limitations are commonly encountered by organizations relying on outdated systems:

- **Scalability constraints** – Physical hardware restrictions make it challenging to scale on-premises systems to meet increased data demands. Scaling often involves significant investments in new hardware, which are both costly and time-intensive, limiting an organization's ability to manage demand spikes effectively.
- **High operational costs** – Maintaining legacy data warehouses involves substantial expenses, including costs for hardware, manual maintenance, and regular system upgrades. As data needs continue to grow, these costs can quickly outpace the value provided by the system.
- **Slow response times and bottlenecks** – Growing data volumes create performance issues in traditional systems, leading to slower queries, limited support for concurrent access, and complex data retrieval processes. These bottlenecks delay insights, impacting the organization's ability to respond swiftly to data-driven needs.
- **Data silos and limited integration** – On-premises systems often operate in isolation, making it difficult to consolidate data across departments. This siloed data limits the scope of analysis and prevents businesses from accessing a full, integrated view of their operations.



DID YOU KNOW?

Organizations that migrate to Azure Databricks from legacy systems often see significant reductions in total cost of ownership, ranging from 30–50%, as they shift from maintaining costly on-premises infrastructure to a fully managed, scalable cloud environment. This allows teams to redirect resources from system management to data analysis and innovation.

The benefits of a modern cloud-native platform

Transitioning to a cloud-native platform like Azure Databricks allows organizations to replace outdated infrastructure with a flexible, scalable environment that supports a broad range of data science and analytics needs.

Key advantages of migrating to Azure Databricks include:

ELASTIC SCALABILITY	Azure Databricks allows independent scaling of compute and storage resources, enabling organizations to meet periods of high demand without heavy hardware investments. This elasticity supports cost-effective scaling, as businesses can adjust based on real-time needs.
AUTOMATED MAINTENANCE AND UPDATES	With Azure Databricks, infrastructure maintenance and software updates are handled within the Azure ecosystem. Automated cluster management and system optimizations allow data teams to focus on analytics and strategy rather than upkeep.
ENHANCED ACCESSIBILITY AND COLLABORATION	As a cloud-based solution, Azure Databricks enables secure and performant data access for global teams, providing shared workspaces that allow data engineers, scientists, and analysts to collaborate in real time. This environment fosters greater alignment and faster insights across the organization.

CASE STUDY 1 (FINANCIAL SERVICES):

Belfius Bank – Enhanced insights for decision-making with Azure Machine Learning

Belfius Bank, a major financial institution in Belgium, utilized Azure Machine Learning to improve its credit risk assessments and deliver real-time financial insights. By transitioning its data operations to the Azure platform, Belfius gained the ability to analyze massive datasets, driving faster and more accurate insights. This transition not only optimized risk analysis but also improved customer experience by delivering personalized banking solutions efficiently.



SECTION 3:

Azure Databricks as a catalyst for transformation

Streamlining data operations

As the first-party solution with Microsoft, Azure Databricks is a powerful transformation tool that goes beyond simply replacing legacy systems. This native integration ensures seamless compatibility with Azure services, enterprise-grade security, and optimal performance within the Azure ecosystem. With a single platform for data engineering, analytics, and machine learning, it enables organizations to move from static, resource-heavy infrastructure to a cloud-native environment that supports integrated data operations and delivers advanced insights.

Azure Databricks brings data engineering, analytics, and AI into one unified environment. This approach minimizes silos, streamlines complex workflows, and enhances collaboration, enabling teams to manage the entire data lifecycle – from ingestion to insight – on a single platform. By centralizing these operations, businesses can access real-time insights and conduct batch processing in ways that are not feasible with legacy systems.

Key features that drive this transformation include:



Real-time and batch processing – Azure Databricks allows for both real-time streaming and batch data processing, letting organizations react quickly to business changes and handle large-scale analytics efficiently.



Scalable AI and machine learning – Azure Databricks facilitates the deployment of AI into production, supporting both classical and generative AI applications. Integrated with widely used machine learning tools like MLflow, PyTorch, and TensorFlow, the platform enables organizations to efficiently manage the full AI lifecycle, from experimentation and model training to deployment. This end-to-end support allows teams to seamlessly operationalize AI across business functions, driving insights and automation at scale.

Accelerating time to insight with real-time processing

Real-time processing capabilities in Azure Databricks allow businesses to act on data as events unfold, providing a strategic advantage in industries where timing is critical. The platform's structured streaming functionality supports real-time monitoring, making it ideal for applications in finance, retail, and manufacturing where rapid response can impact outcomes significantly.

For example:



Streaming analytics for live insights – With structured streaming and Delta Live Tables (DLT), Azure Databricks enables continuous monitoring and analysis, allowing teams to respond to customer needs, operational shifts, and market changes in real time. DLT automates the setup and maintenance of reliable, high-quality data pipelines, ensuring data accuracy and consistency for real-time analytics.



Low-latency processing – By reducing latency from data collection to insight, Azure Databricks supports fast decision-making for applications like fraud detection, supply chain optimization, and customer experience management.

Driving innovation with AI and machine learning

Azure Databricks provides a comprehensive toolkit for AI and machine learning, designed to empower organizations to innovate with minimal complexity:



End-to-end MLOps and AI support – Azure Databricks provides comprehensive MLOps capabilities that extend beyond traditional lifecycle management. With MLflow for experiment tracking and deployment, as well as integrated support for Vector Search, GPU model serving, and hosted LLMs, the platform helps organizations put both classical and generative AI into production. Additional tools like the feature store, agent framework for autonomous workflows, and model evaluation and fine-tuning capabilities ensure seamless scaling and governance for AI applications.



Comprehensive model deployment and monitoring – The platform's robust model management features, including model monitoring, AutoML for rapid prototyping, and support for continuous improvement, enable teams to deploy and monitor AI models effectively. Azure Databricks also offers advanced capabilities like the feature store and model tracking across diverse use cases, driving AI and ML insights with minimal manual intervention.

Advantages of the Azure Databricks platform

Azure Databricks stands out for its seamless integration within the Azure ecosystem and flexible scalability. These features make it an optimal choice for organizations looking to modernize their data capabilities:

- **Deep integration with Azure services** – As a native part of Azure, Azure Databricks integrates smoothly with other Azure tools like Azure Data Lake Storage, Azure OpenAI, and Azure Machine Learning, creating a streamlined data pipeline from ingestion to advanced analytics without requiring complex setups.
- **Elastic scalability** – Unlike fixed-capacity legacy systems, Azure Databricks allows organizations to scale compute and storage resources independently, helping them adapt to workload changes efficiently while managing costs.
- **Data governance and compliance** – The built-in Unity Catalog provides centralized data governance with access control, usage tracking, and data lineage capabilities, helping organizations meet regulatory requirements and maintain data integrity across their operations.



CASE STUDY 2 (BANKING):

Trusting Social – Real-time credit scoring in Vietnam

Trusting Social, a fintech company in Vietnam, leveraged Azure to build a real-time credit scoring system, enabling accurate financial risk assessment for previously underserved populations. With Azure's machine learning and data processing capabilities, Trusting Social provides quick, reliable credit scores, promoting financial inclusion and enhancing decision-making within Vietnam's dynamic lending environment.

SECTION 4:

The strategic advantage of migrating to Azure Databricks — AI and real-time analytics

Maximizing AI and ML capabilities

Azure Databricks optimizes the entire machine learning process lifecycle, making it easier for organizations to integrate predictive and prescriptive insights across their operations.

Key features include:



End-to-end ML lifecycle management – MLflow, the integrated ML management tool, supports the complete ML lifecycle. From experimentation to deployment, MLflow enables teams to track, manage, and control models efficiently in production environments.



Seamless data access for ML – Direct access to large datasets through Azure Data Lake Storage and Azure Synapse Analytics enables robust model training, allowing data scientists to work with complete, real-time data without cumbersome data transfers.



Automated machine learning (AutoML) – With AutoML, users can quickly build, tune, and select models without extensive manual effort. This is especially useful for teams that want to accelerate ML experimentation without needing deep in-house expertise.

Real-time analytics for rapid insights

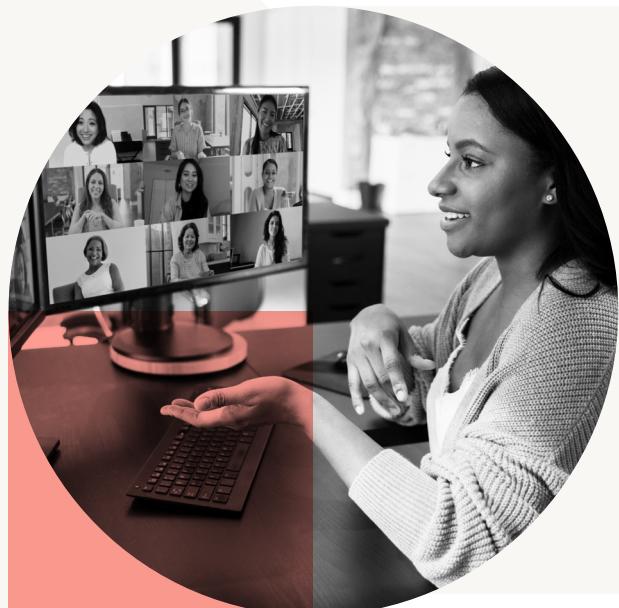
In fast-paced industries, real-time data processing allows businesses to respond immediately to changing conditions. Azure Databricks supports real-time analytics that can transform operational efficiency and responsiveness:

- **Structured streaming for live data** – By processing data as it arrives, structured streaming in Azure Databricks is ideal for real-time applications such as fraud detection, inventory management, and customer engagement. Businesses gain a continuous view of operations, enabling prompt actions.
- **Low-latency processing** – With minimized latency, Azure Databricks helps companies transition from data ingestion to actionable insights quickly. This rapid processing is vital for sectors like finance and retail, where quick, informed decisions can impact profitability and customer satisfaction.

Scaling data operations to drive innovation

Azure Databricks supports scalable data operations, making it easy for organizations to run complex analytics and machine learning models across large datasets:

- **Data engineering at scale** – Leveraging Apache Spark's distributed computing, Azure Databricks enables efficient handling of large data volumes and complex transformations, supporting use cases from routine Extract Transform Load/Extract Load Transform (ETL/ELT) to advanced data science.
- **Predictive analytics for strategic decision-making** – By applying machine learning to forecast trends, predict behavior, and optimize resources, Azure Databricks transforms raw data into actionable, forward-looking insights that inform decisions in marketing, operations, and beyond.



CASE STUDY 3 (MEDIA):

MultiChoice – Streamlining content creation with Azure

MultiChoice, a leading entertainment company in South Africa, used Azure to consolidate its data, streamline content creation, and improve distribution efficiency. By moving its operations to Azure, MultiChoice could leverage data insights to personalize viewer experiences, optimize content recommendations, and simplify data management across global teams. Azure's scalable solutions have empowered MultiChoice to better serve its diverse audience with relevant, engaging content.

Enhanced value through Azure integrations

A key benefit of Azure Databricks is its seamless integration within the Microsoft Azure ecosystem, allowing organizations to unify their analytics, machine learning, and data governance across services. Azure Databricks combines both built-in Azure technologies and robust integration capabilities for enhanced functionality:



Built on Microsoft's Azure stack – As a native Azure service, Azure Databricks offers tight integration with core Azure components, ensuring compatibility with Azure's security, compliance, and scalability features. Azure Databricks also integrates with Azure Fabric, Microsoft's unified data platform that brings together data integration, analytics, and governance. This first-party service facilitates easier setup and unified management for data and AI workflows within the Azure environment, providing enhanced flexibility and operational efficiency.



Unified integrations with Azure services – Azure Databricks connects with various Azure services, including Azure Machine Learning for full ML lifecycle management, Azure DevOps for streamlined Continuous Integration/Continuous Deployment (CI/CD), Key Vault for secure credential management, and private links for network security.



Leverage advanced generative AI – Azure Databricks supports Azure OpenAI and advanced integrations with LLMs for generative AI capabilities, delivering versatile options for both classical ML and generative AI workflows.

These integrations and built-in features help organizations create end-to-end data and AI solutions, transforming data into a consistent, strategic driver for growth and innovation.

SECTION 5:

Key benefits of migrating to Azure Databricks

Migrating to Azure Databricks opens new possibilities for organizations looking to enhance efficiency, reduce costs, and leverage data-driven insights. This cloud-native platform offers unmatched scalability, flexibility, and advanced analytics, far surpassing the limitations of legacy systems.

Reducing operational overhead

Azure Databricks significantly reduces operational burden by handling infrastructure, software updates, and system maintenance. This managed approach allows organizations to shift resources from routine maintenance to strategic data initiatives, focusing on analytics, machine learning, and innovation. Furthermore, Azure Databricks' automated scaling adjusts resources dynamically based on workload demand, ensuring peak performance during high-demand periods without over-provisioning. With its integrated tools for data engineering, analytics, and machine learning, Azure Databricks also reduces the need to transfer data across multiple systems, enhancing workflow efficiency and ensuring data accessibility and consistency.

Enhancing scalability and flexibility

Legacy systems often struggle to adapt to expanding data volumes and complex analytic demands. In contrast, Azure Databricks enables organizations to scale compute and storage resources independently, maintaining cost efficiency while meeting evolving workload requirements. The platform's support for both real-time streaming and batch processing allows organizations to handle diverse use cases on a single platform, making it versatile enough for immediate insights as well as large-scale data transformations. As data needs grow, Azure Databricks offers elasticity to accommodate future expansion without costly hardware upgrades, supporting seamless scaling over time.



Leveraging advanced analytics and AI

Azure Databricks empowers organizations to advance from basic analytics to more sophisticated AI and machine learning applications. By integrating with MLflow, the platform simplifies the machine learning lifecycle, including model tracking, versioning, and deployment, ensuring efficient and consistent ML operations. AutoML further accelerates model development by automating the testing and optimization of multiple models, making it possible for teams to deploy AI solutions without extensive data science expertise. In addition, Azure Databricks supports real-time AI applications, enabling organizations to act immediately on new data for use cases such as customer engagement, fraud detection, and dynamic pricing.

Improved data governance with Unity Catalog

Data governance is essential for organizations handling sensitive or regulated data, and Azure Databricks addresses this need with Unity Catalog. This centralized governance solution provides a unified control layer for managing data permissions, tracking usage, and ensuring data lineage across workflows. Access control enables organizations to set permissions aligned with internal policies, safeguarding sensitive data through controlled access. With data lineage tracking, Unity Catalog allows organizations to trace data origins and usage, providing transparency for regulatory compliance and ensuring comprehensive data governance.

Cost savings and predictable pricing

Azure Databricks operates on a consumption-based model, allowing organizations to pay only for the resources they use. This accommodates variable workloads, eliminating the need for upfront costs and helping control expenses. The managed infrastructure reduces the total cost of ownership by minimizing expenses associated with on-premises maintenance, hardware upgrades, and system monitoring, enabling teams to reallocate budgets toward high-value initiatives like analytics and AI development. Integrated cost management tools within Azure allow organizations to monitor usage, set budget alerts, and optimize resources, ensuring effective cost management and supporting teams in staying within budget.

SECTION 6:

Enhancing data governance and security with Unity Catalog

As data governance and security requirements grow, Azure Databricks' Unity Catalog offers an effective solution for centralized management, compliance, and data protection. Unity Catalog simplifies governance processes, ensures consistent policy enforcement, and integrates with Azure's security features to protect sensitive information.

Centralized governance and data access control

Unity Catalog centralizes governance with a unified interface for managing access controls, tracking data usage, and monitoring data lineage across departments:

-  **Single governance layer** – Manage permissions and policies for all data assets from one place, enabling efficient oversight and consistent enforcement across Azure Databricks.
-  **Access control** – Define access policies based on user roles and responsibilities to ensure that only authorized users access sensitive data.
-  **Granular permissions** – Assign permissions at table, column, or row levels, allowing precise control over who can access specific data assets.

Data lineage for transparency and compliance

Unity Catalog's data lineage features provide visibility into data origins, transformations, and usage, supporting compliance and audit readiness:

-  **Transparency into data transformations** – Track each stage of data processing from ingestion to analysis, creating an audit trail for accuracy and transparency.
-  **Compliance support** – Generate reports detailing data movement and transformations, meeting regulatory standards like GDPR and CCPA.

Simplified compliance with automated governance

Unity Catalog automates key aspects of data governance, making it easier to enforce policies and manage data securely:

-  **Automated data tagging and classification** – Tag data assets by sensitivity or business function, applying automatic access controls to safeguard sensitive information.
-  **Consistent policy enforcement** – Automated governance ensures data policies apply uniformly across assets, reducing security gaps and supporting compliance as data grows.

Enhanced security with Azure integration

Unity Catalog seamlessly integrates with Azure's security frameworks to enhance data protection:

-  **Azure Active Directory (AAD) integration** – Centralize user authentication and apply AAD policies across Azure Databricks for a secure, unified login experience.
-  **Data encryption** – Secure data at rest and in transit, combining encryption with Unity Catalog's governance controls to protect sensitive information.
-  **Audit logs and monitoring** – Azure's logging tools track user actions and data access, enabling security teams to monitor activity and detect unusual behavior.
-  **Integration with Azure Purview** – Gain enterprise-wide data visibility and governance by integrating with Azure Purview, which gives organizations a comprehensive view of their data estate. This helps support effective data cataloging, lineage tracking, and compliance management.

Supporting long-term governance goals

Unity Catalog helps build a sustainable governance framework adaptable to changing needs:

ROUTINE GOVERNANCE AUDITS	TRAINING ON GOVERNANCE BEST PRACTICES	ADAPTABLE POLICIES
Use lineage tracking and Role-Based Access Control (RBAC) for regular audits, ensuring compliance with evolving regulations.	Educate teams on handling sensitive data using Unity Catalog's classification and tagging features.	Scale governance policies with organizational growth, ensuring data access and compliance keep pace with business demands.

SECTION 7:

Preparing for a successful migration to Azure Databricks

Migrating to Azure Databricks is a strategic move that requires careful planning to ensure a seamless transition from legacy systems. This section provides a step-by-step guide to preparing for migration, from initial assessment to execution, to help organizations maximize the value of Azure Databricks.



Step 1: Assess your current data infrastructure and needs

Begin by evaluating your existing data environment to understand potential challenges and needs for migration.

- **Inventory data sources and formats** – Document all data sources, including transactional databases, streaming data, and unstructured data. Understanding data formats will facilitate a smoother transition. Tools like Impetus' LeapLogic automate the assessment of diverse workloads, providing comprehensive insights into data assets and their complexities.
- **Identify application dependencies** – Map out dependencies on legacy systems to prevent disruptions during migration. Again, LeapLogic offers detailed analysis of workloads and their interdependencies, enabling organizations to visualize complex relationships and plan migrations effectively.
- **Define data processing requirements** – Clarify performance needs, such as real-time analytics or machine learning, to configure Azure Databricks for optimal processing.



Step 2: Develop a phased migration strategy

A phased approach helps manage risk, allowing teams to test and optimize each step.

- **Prioritize workloads** – Start with non-critical workloads to test compatibility and performance in Azure Databricks. Move critical workloads only after initial success.
- **Set timelines and allocate resources** – Outline a realistic timeline and assign resources for each phase, ensuring that necessary personnel are available for key stages.
- **Plan resource allocation** – Azure Databricks' flexible scaling allows temporary resource expansion during migration, speeding up data transfers and processing.



Step 3: Leverage Azure Databricks migration tools and support

Azure Databricks offers various tools and resources to streamline migration and ensure data integrity.

- **Leverage data migration services** – Microsoft and Databricks provide migration services that help simplify the process and ensure best practices. This includes partner tools such as Fivetran and Qlik.
- **Monitor data flows with Delta Lake** – Delta Lake's change data capture and rollback features allow teams to track data accuracy in real time.
- **Consult Databricks Academy** – Training through Databricks Academy equips data teams with the necessary skills, maximizing the platform's benefits.



Step 4: Prepare data for migration

Ensure data compatibility and integrity before migration by making necessary adjustments.

- **Clean and validate data** – Remove duplicate entries and inconsistencies to enhance data quality in the new environment.
- **Standardize formats and schemas** – Align data schemas with Azure Databricks standards to simplify integration and future workflows.
- **Enable Delta Lake** – Use Delta Lake to maintain data consistency with Atomicity, Consistency, Isolation, Durability (ACID) transactions and data versioning, supporting reliable data handling throughout migration.



Step 5: Test and validate migrated data

Comprehensive testing ensures that data integrity, workflows, and performance meet operational standards.

- **Validate data accuracy** – Cross-check migrated data against original sources to confirm accuracy and completeness.
- **Run performance benchmarks** – Test analytics workflows to confirm that Azure Databricks meets scalability and speed requirements, conducting load tests for high-demand scenarios.
- **Conduct user acceptance testing (UAT)** – Involve end users to ensure that the migrated system meets their needs and identify any usability issues before going live.



Step 6: Facilitate change management and team training

Successful migration extends beyond technical deployment; it requires engaging and equipping the people who will use the system.

- **Engage stakeholders early** – Include data engineers, scientists, and analysts throughout the migration process to align with business goals and address user needs.
- **Provide comprehensive training** – Offer structured training on data management, analytics, and machine learning in Azure Databricks, tailored to each team's needs.
- **Develop support resources** – Create documentation, troubleshooting guides, and best practices to empower teams post-migration, minimizing operational disruptions.

SECTION 8:

Measuring success and maximizing value after migration

Migrating to Azure Databricks is just the start. To ensure sustained success, organizations should measure key outcomes, optimize workflows, and leverage Azure Databricks' full potential. Establishing key performance indicators (KPIs), enhancing operations, and integrating Azure services all contribute to maximizing value post-migration.

Establishing KPIs to measure migration success

Setting clear KPIs allows organizations to quantify the impact of migration and ensure alignment with business goals. Key areas include:

- **Operational efficiency** – Track metrics like data processing speeds and system uptime to measure improvements in efficiency and data flow.
- **Cost reductions** – Compare pre- and post-migration expenses to assess cost savings, such as reduced hardware, maintenance, and resource costs.
- **Data accessibility and insights** – Monitor query performance and user access to evaluate improvements in data availability and speed of insights.



CASE STUDY 4 (MANUFACTURING):

Parker Hannifin – Improving data processing in manufacturing workflows

Parker Hannifin, an industrial manufacturing leader in the U.S., adopted Azure Databricks to enhance data processing and support innovation across its operations. The Azure platform enabled Parker Hannifin to streamline its data engineering tasks, reduce manual intervention, and automate production insights, improving operational efficiency and enabling data-driven decisions in real time. Azure Databricks helped Parker Hannifin stay competitive in a demanding manufacturing sector.

Optimizing workflows and operations on Azure Databricks

After migration, refining workflows enhances efficiency and cost-effectiveness, unlocking more of Azure Databricks' capabilities.

- **Leverage Delta Lake** – Implement Delta Lake for ACID transactions and data versioning, improving reliability for applications that need consistent data.
- **Enable autoscaling** – Use autoscaling to adjust resources based on demand, reducing costs while ensuring performance remains stable.
- **Optimize data layout** – Efficient data layout with liquid clustering and predictive optimization speed up queries for large datasets, enhancing user experience and accessibility.

Driving innovation with AI and advanced analytics

Azure Databricks provides a solid foundation for advanced analytics, enabling organizations to expand beyond standard reporting into predictive insights and AI-driven processes.

- **Expand ML initiatives** – Use MLflow for tracking and managing models, fostering consistent experimentation and streamlined model versioning.
- **Automate with AutoML** – AutoML automates model building and validation, making it easier to scale analytics without extensive data science resources.
- **Implement real-time analytics** – Structured streaming supports live data analysis, allowing for use cases such as customer engagement and operational monitoring.

CASE STUDY 5 (FINANCIAL SERVICES):

Deutsche Börse – Enhanced analytics in financial operations

Deutsche Börse, a German stock exchange, enhanced its financial analytics by utilizing Azure Databricks to process large-scale data in real time. This transition enabled the exchange to optimize trading operations, reduce latency, and provide investors with timely, data-driven insights. Azure Databricks' scalable infrastructure supported Deutsche Börse's need for fast, reliable analytics, enhancing operational efficiency and customer satisfaction.



Maximizing ROI with Azure integrations

Extend the value of Azure Databricks by integrating it with other Azure services for a holistic data strategy.

- **Azure Machine Learning** – Combine with Azure Machine Learning for seamless model deployment and monitoring, supporting the full ML lifecycle.
- **Power BI** – Unlock powerful analytics and visualization capabilities with direct publishing of aggregated data to Power BI, allowing faster access to insights and enhanced data-driven decision-making. This integration supports real-time data exploration, making it possible to build dynamic, interactive dashboards that drive business intelligence across the organization.
- **Azure Fabric** – Seamlessly integrate with Azure Fabric to enhance collaboration across teams and simplify the management of complex data estates while leveraging the full Azure ecosystem for scalable, secure, and efficient data operations.

Continuous improvement for long-term value

Adopting continuous improvement practices ensures Azure Databricks continues to meet evolving business needs.

REFINE GOVERNANCE POLICIES	MONITOR RESOURCE USAGE	PROMOTE KNOWLEDGE SHARING
Regularly update Unity Catalog's governance policies to keep data secure and compliant as usage grows.	Use Azure's cost management tools to track resource consumption, optimizing allocations to control costs.	Encourage cross-team collaboration and best practice sharing to stay aligned on updates and maximize Azure Databricks' potential.

SECTION 9:

Future-proofing your data strategy with Azure Databricks

To remain competitive, organizations need a data strategy that adapts to future demands. Azure Databricks provides a flexible and scalable platform that supports current needs while adapting to the evolving landscape of data analytics, AI, and compliance. With capabilities for seamless integration, advanced analytics, and robust data governance, Azure Databricks helps organizations build a future-ready data foundation.

Building a scalable foundation for long-term growth

Azure Databricks offers the scalability needed to meet expanding data requirements. With dynamic compute and storage scaling, organizations can respond to changes in data processing demands without costly infrastructure overhauls.

- **Elastic compute and storage scaling –** Azure Databricks allows compute and storage resources to scale independently based on demand, ensuring efficient use of resources. This elasticity supports workload spikes and enables organizations to scale seamlessly as data needs grow.
- **Optimized for variable workloads –** Azure Databricks handles both real-time streaming and batch processing, making it ideal for managing diverse data workflows. This flexibility enables organizations to balance day-to-day data processing with large-scale analytics, supporting varied business needs.

Adapting to emerging technologies and analytics trends

Azure Databricks provides flexibility and forward compatibility through its deep integration with open-source tools and regular feature updates, allowing organizations to incorporate new analytics technologies as they emerge.

- **Open-source tool compatibility** – Built on a foundation of open-source standards like Apache Spark, Delta Lake, and MLflow, Azure Databricks is compatible with widely used data science libraries such as TensorFlow, PyTorch, and Apache Iceberg. This robust, open-source foundation allows organizations to integrate evolving analytics frameworks, ensuring a flexible, interoperable environment without disrupting existing workflows.
- **Continuous platform innovations** – As a fully managed cloud service, Azure Databricks benefits from regular updates that bring the latest in data processing and analytics advancements. Key updates include vector search for generative AI applications, GPU model serving for high-performance model deployment, and advanced tools for model evaluation and fine-tuning. These updates ensure that businesses have access to cutting-edge capabilities, enabling them to stay competitive with minimal manual configuration.

Driving innovation with AI and machine learning

Azure Databricks is optimized for AI and machine learning, providing a foundation that supports predictive analytics, automation, and intelligent decision-making.

MACHINE LEARNING LIFECYCLE MANAGEMENT	Azure Databricks offers end-to-end support for the ML lifecycle through MLflow, enabling efficient tracking, versioning, and deployment of models. This structured ML workflow supports growth and adapts to evolving data science practices.
AUTOMATED MACHINE LEARNING (AUTOML)	AutoML on Azure Databricks accelerates model development by automatically testing and selecting optimal models. This tool enables organizations to quickly scale machine learning capabilities, helping teams keep pace with expanding AI demands.

Reinforcing data governance for a secure future

Azure Databricks includes Unity Catalog, a comprehensive governance solution that scales with data environments to ensure long-term data security and compliance.

- **Adaptable governance with Unity Catalog** – Unity Catalog provides a centralized governance framework with fine-grained access controls and data lineage tracking. This system supports growing data volumes and changing compliance needs, allowing organizations to scale governance practices seamlessly.
- **Proactive compliance and audit readiness** – Unity Catalog's governance features simplify compliance with data privacy regulations, such as GDPR and CCPA. Data lineage tracking and role-based access control (RBAC) ensure regulatory standards are met, helping organizations stay prepared for future audits.

Building security with Azure integrations

By integrating with Azure's enterprise security features, Azure Databricks helps organizations maintain a secure environment as they scale.

- **Identity and access management with AAD** – Unity Catalog integrates with AAD to centralize user authentication, ensuring a secure, unified login experience across Azure services.
- **Encryption for data protection** – Data encryption in Azure Databricks safeguards information both in transit and at rest, supporting data integrity and security across the platform.
- **Audit logs and monitoring** – Azure's built-in audit and monitoring tools allow organizations to track access events and data modifications, providing security oversight that grows with data usage.

SECTION 10:

Driving growth with Azure Databricks

Azure Databricks delivers a powerful and flexible data platform that helps organizations drive immediate growth. By migrating from legacy systems, businesses can achieve faster insights, reduce infrastructure costs, and focus on data-driven initiatives that enhance customer experience and operational efficiency. Azure Databricks empowers data teams to work in a unified environment, accelerating innovation and supporting strategic objectives for lasting growth.

- **Accelerated decision-making** – Azure Databricks supports real-time analytics, enabling organizations to process data as it is generated and respond swiftly to new information. This capability is vital for data-driven decisions that boost operational efficiency and responsiveness, especially in fast-paced industries like finance and retail.
- **Cost savings and operational efficiency** – Migrating to Azure Databricks reduces the need for costly on-premises maintenance, eliminates hardware expenses, and enables optimized resource usage through autoscaling. The platform's managed environment allows organizations to allocate resources according to demand, resulting in significant operational savings.
- **Unified data and AI workflows** – By consolidating data engineering, analytics, and machine learning on a single platform, Azure Databricks minimizes the need for data transfers across multiple systems, ensuring faster insights and simplifying the data-to-insights journey. Teams can work collaboratively on data and AI workflows, accelerating innovation and driving quicker results that support growth.

- **Enhanced customer experience through data-driven insights** – Azure Databricks enables advanced analytics and personalized recommendations, helping organizations deliver superior customer experiences. From real-time fraud detection to predictive maintenance and targeted marketing, the platform supports applications that directly contribute to customer satisfaction and business expansion.
- **Growth-driven AI capabilities** – With support for both classical and generative AI, Azure Databricks offers tools to develop, deploy, and monitor AI models at scale. Features like the feature store, model monitoring, and integration with Azure Machine Learning enable organizations to build robust AI-driven processes that adapt as customer needs evolve, further driving growth.

Building a resilient, future-ready data environment

Azure Databricks offers a foundation that allows organizations to stay current with advancements in data and analytics. By leveraging Azure Databricks' continuous updates and seamless integrations with other Azure services, businesses can constantly adapt to new technologies and industry trends.

With Azure Databricks, companies are prepared to scale their data operations, handle increasing data volumes, and adopt emerging analytics tools — all while ensuring that data remains secure and governed.

Taking the next step

For organizations ready to modernize their data environment, Azure Databricks offers a complimentary assessment through the Analytics Modernization Accelerator (AMA) program. This initiative provides expert guidance and resources to facilitate a seamless transition to Azure Databricks, helping organizations quickly realize the benefits of a cloud-native data platform.

By leveraging this program, businesses can gain a competitive edge and set the stage for data-driven innovation. With a clear migration plan, well-defined KPIs, and a culture of continuous improvement, organizations can maximize their investment in Azure Databricks and unlock the full potential of their data assets. The platform empowers data teams to focus on what matters most — driving insights, improving customer experiences, and achieving strategic objectives.

Azure Databricks transforms how organizations manage and leverage data, enabling them to respond to changing demands and embrace new opportunities. By migrating to a flexible, scalable platform, companies can make data a core part of their growth strategy and achieve long-term resilience in an increasingly data-centric world.

To learn more about the AMA Program and initiate your free assessment, contact your Databricks or Microsoft Sales Representative.

[Sign up for a free trial](#)

