databricks ✕ **Microsoft**

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.
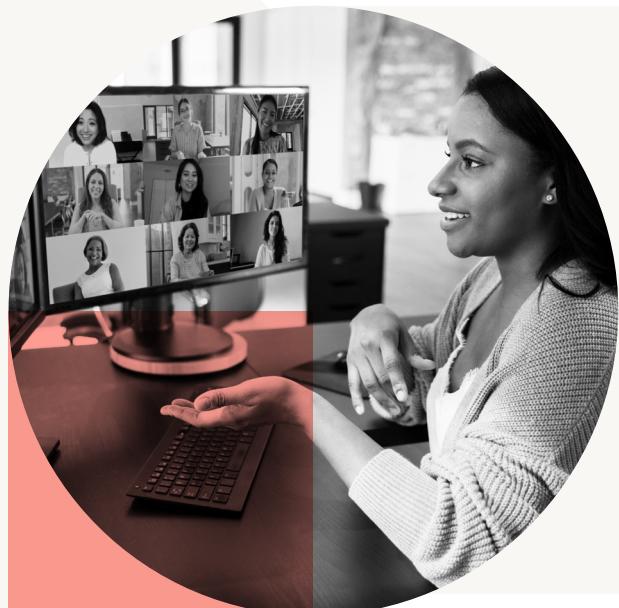
Real-time analytics for rapid insights

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.

databricks ✕ Microsoft

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.