databricks

The Guide to Data and AI Transformation at Scale

The first step to increased business value from sensitive data with confidential computing

Microsoft AMD

databricks

Reimagining the value of data

You've probably heard data called "the new oil," but that analogy vastly understates its value and potential.

The importance of data is closer to that of water than oil to modern business. In its native form, it's vital. And more than that, it can be transformed to meet a host of needs. Data is essential to business survival, competitive differentiation, and innovation.

So when data is impeded, trapped in silos, slowed by the inability to scale, or tainted by noncompliance with regulations or by breaches, its value can plummet. It's a common issue, with 87% of organizations[1] facing challenges to deliver on their data strategy. These challenges include the deployment of modern data architectures, the secure and efficient leveraging of data, compliance with ever-changing regulations, resourcing the right talent, and identifying and executing on AI opportunities. To overcome these challenges, a new strategy for storing, accessing, curating, and gaining value from data must be found.

An effective data strategy must be carefully considered and must solve for the difficult issues of working with sensitive and confidential data. This can increase the value of data in use cases that range from personal financial data to cancer screening, sales transaction histories, and beyond. The need to both secure and use these types of information is critical for ongoing business success. In the coming years, Microsoft Azure confidential computing will play a central role as the market for confidential computing expands at a projected annual rate of 94.4% by 2026.[2] To benefit from that immense growth, organizations will need a solution to:

- Unlock the potential of sensitive data, analytics, and AI for more innovation and business value
- Minimize risk and simplify governance with a unified governance model across clouds
- Lower operational costs by reducing legacy processes
- Increase data confidentiality and privacy while sending, storing, and processing sensitive data

In this eBook, we'll describe the first stage to a transformation that enables all these at scale: Identifying the goals and value of a comprehensive and executable strategy as a North Star for decision-making that's powered and secured using Azure Databricks, Azure confidential computing, and AMD Secure Encrypted Virtualization-Secure Nested Paging (SEV-SNP) technology.

[1] "Building a High-Performance Data and AI Organization," MIT Technology Review Insights in partnership with Databricks, 2021.
[2] "Confidential Computing Market Size, Share, and COVID-19 Impact Analysis," Fortune Business Insights, July 2023.

Microsoft AMD

databricks

Do you know what your priority goals are?

Most organizations on a data, analytics, and AI journey establish a set of priority goals for their sensitive and confidential data. The goals generally fall into one of four categories:

1. Business outcomes
Legacy systems can make it difficult to adapt quickly to dynamic market opportunities and emerging risks. Digital transformation can be an opportunity to build a new technology foundation to increase business value with more agility, scalability, security, and ease.

2. People
People are your most valuable asset, and the battle for top talent is fiercer than ever. The goal is to give your people a frictionless data environment so they can access the information they need to do their best work, spending less time collecting and compiling data and more time developing insights and algorithms.

3. Technology
Complex system architectures, vendor lock-in, and proprietary solutions are expensive, ineffective, and slow to evolve. The rise of cloud computing has made the move from investing in equipment (capital expenditure, or CapEx) to investing in capabilities (operating expense, or OpEx) a necessity for scale and innovation. But moving to a cloud-based solution is not sufficient. It must be done in a way in which data security is paramount. Azure confidential computing, with its isolated trusted execution environments (TEEs) powered by AMD SEV-SNP technology, is a highly effective way to unlock the value of sensitive data more securely.

4. Privacy and security
As your most valued business asset, your data must be both tightly secured and available for innovation to derive its full value. By employing advanced security measures, such as in-memory encryption and verification of the underlying cloud environment, confidential computing enables organizations to leverage most sensitive data in the cloud, reducing security risk when data is being sent, stored, or used in memory.

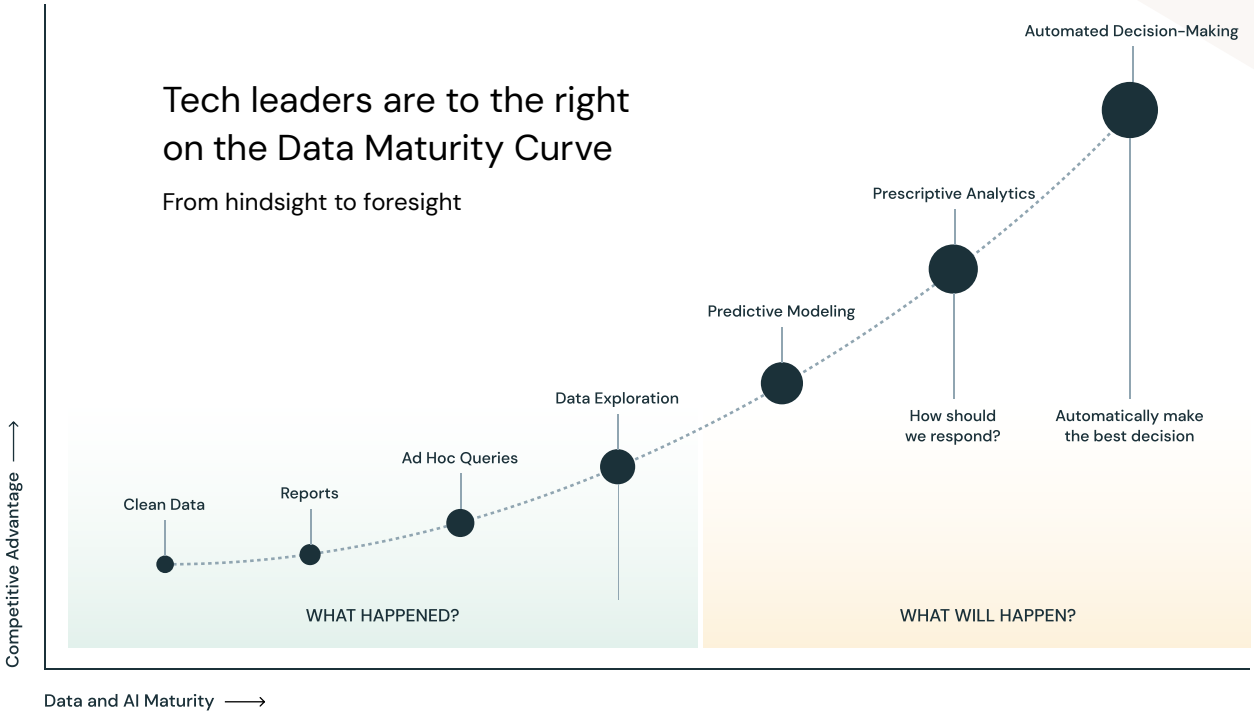
To get buy-in, gain C-level clarity
Successful transformation for data, analytics, and AI needs to be a company-wide initiative with leadership support. From the start, make sure your C-suite is clear and united on how the business, the people, and the technology benefit from your initiative — even if you start with a pilot in a single business unit.

databricks

Before you start: See the barriers

The journey from a data-driven organization to a data and AI-driven one is a true transformation — not just an evolution — of your organization's data usage. Before you start, it's important to know what barriers stand in your way. Here are the seven most prevalent:

1. Uncertainty of AI maturity. Pockets of success in analytics and AI within the organization don't move you up the maturity curve, and replicating and scaling success is nearly impossible. Use the Data Maturity Curve below as a guide to take a company-wide look at where you stand.



2. Exploding data volumes and types. It's not only volume but data types — unstructured video, audio, narrative text — that have outgrown even the most modern approaches to SQL-based data processing.

3. The fog of insights. These large data volumes make it nearly impossible to state, by priority, how data insights can be achieved or how the business should react to changing data.

4. High data-processing costs. You simply cannot afford to hire the number of people needed to respond to every piece of data flowing into your environment. Machines scale; people do not.

5. Risk of sensitive data disclosure in the cloud. Data needs to be secured in all its states while in transit, in storage in the cloud, and while it's being processed. While there are methods in the cloud for addressing the first two, deploying Azure Databricks clusters on AMD SEV-SNP confidential virtual machines (VMs) helps mitigate the risk of disclosure of sensitive data while it's in use in memory.

6. Compliance requirements for handling sensitive data. Safe handling of personal data required by industry and government regulations has made many organizations wary about migrating this type of data to the cloud. But some of the most valuable business insights can be acquired by unlocking access to this type of data.

7. Traditional processing and data management won't work. Without the enhancements of machine learning and AI, your workforce will never be able to leverage modern data volumes to help your organization compete and succeed.

Let's look at the seven solutions to these issues.

databricks

1. Embrace Zero Trust using confidential computing

Organizations seeking to move their most sensitive data to the cloud want to trust their cloud service provider (CSP) as little as possible. Running a Databricks cluster on Azure Databricks helps protect your data during its entire lifecycle: while at rest on disk using customer-managed keys, while in transit on the network using TSL or HTTPS, and now, while in use in memory using Azure confidential VMs with AMD SEV-SNP technology.

These confidential VMs run computations in a hardware-based TEE with memory encryption keys managed by the CPU firmware and inaccessible to Azure operators. Only the VM that writes to memory can read that memory or write over it. This takes the Azure hypervisor and host operating system out of the trusted computing base and allows users of confidential VMs to assume these software layers have been breached. This is just one example of Azure's larger commitment to Zero Trust and defense in depth, which also includes trusted launch, Microsoft Azure Attestation, and Azure Managed Hardware Security Module (HSM).

INSIGHT

Ensure you have maximum control over your sensitive data by protecting it during its entire lifecycle: while at rest on disk, while in transit on the network, and now, while in use in memory using Azure confidential VMs with AMD SEV-SNP technology.

2. Plan for both migration and modernization

In your move to the cloud, avoid a simple "lift and shift" approach: modernize first.

Most on-premises applications are not built with the cloud in mind — and they don't offer the same power or innovative capabilities. In contrast, modern cloud applications are modular in design with RESTful web services for better performance and APIs to easily provide access to an application's functionality.

As a first step toward taking advantage of these and many other cloud benefits, take an inventory of your business-critical applications, prioritize them based on their business impact, and modernize them in a consistent manner for cloud-based deployments. They'll generate and store a significant amount of your data.

INSIGHT

Use a consistent approach to cloud-based application design to make it easier to extract data when it's needed.

INSIGHT

Enable the best customer experiences through real-time execution combined with historical data.

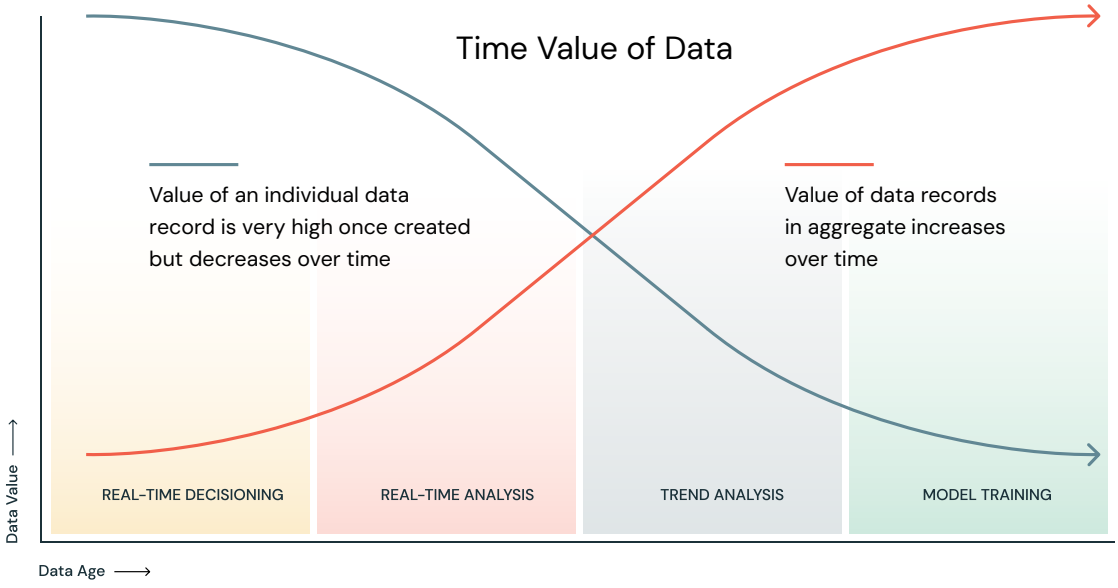
3. Realize the value of both aggregate and real-time data

View the value of data through two different lenses: in the aggregate and as individual, real-time events. Both are necessary in different contexts.

Aggregate data becomes more valuable over time. It allows you to look back and see the complete history of an aspect of your business and to discover trends.

Real-time data is most valuable the moment it's captured and loses value over time. It allows you to act immediately and make better decisions in the moment to reduce risk, provide better service to your customers, lower your operating costs, and more.

For example, credit-card fraud models can use an aggregated view of deep historical data about a given customer's buying patterns (for example, location, day of week, time of day, retailer, average purchase amount) to build rich models that are then executed for each new individual credit-card transaction in real time.



4. Land all data in a data lake

The overall goal for business data is to gain more insights from that data as quickly as possible. That means getting it out of silos. To be a data-, analytics-, and AI-driven organization, you need to be able to store and process all data — regardless of size, shape, or speed — where the people who need it can get it, trust it, and use it right away.

That's where the unique **Databricks Lakehouse Platform** comes in. It's the only solution that unites the reliability, governance, and performance of a data warehouse with the openness, flexibility, and machine-learning support of data lakes.

INSIGHT

Break down silos and accelerate access and insights with the Databricks Lakehouse Platform.

databricks

Tour the Lakehouse

Access the power of Azure Databricks

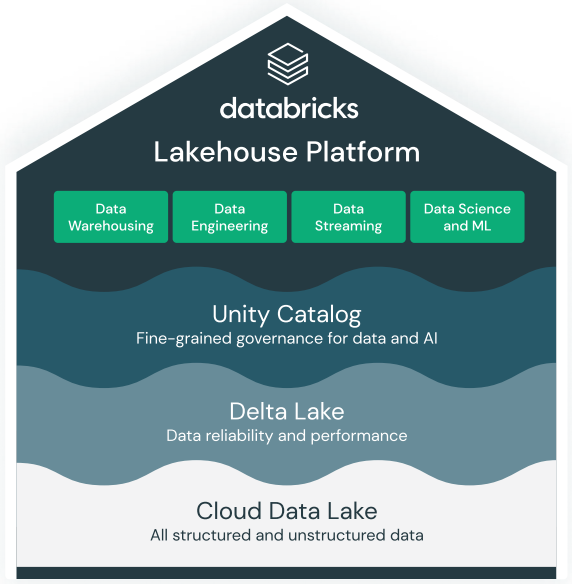
The first step to increased business value from sensitive data with confidential computing

Derive greater innovation and business value from data analytics and AI.
Unlock your data, analytics, and AI potential to modernize your data platform and drive greater innovation and business value.

Reduce risk and simplify governance.
Minimize risk and ensure compliance with a single, consistent governance model. Simplify how you discover, access, and share real-time data. On Azure Databricks, protect data in use in memory for added confidentiality and privacy backed by Azure confidential computing and anchored on AMD Infinity Guard featuring SEV-SNP technology.

Lower total cost of ownership.
Lower costs by reducing legacy processes and sending more workloads to Azure Databricks for an open, scalable, unified cloud data architecture that protects data in memory.

Protect your most sensitive data at rest, in transit, and while in use.
Increase data confidentiality and privacy while collaborating, sending, storing, and processing sensitive data with security features such as customer managed keys for data on disk and hardware-based memory encryption for data in use, powered by Azure confidential computing, featuring AMD EPYC™ CPUs provided by AMD Infinity Guard featuring SEV-SNP technology.



Microsoft AMD

5. Put the user at the center of your efforts

To deepen the impact of your data and AI transformation, focus on improving the user experience — how individuals and groups interact with and get what they need from data. Data has to be easily discoverable, with default access to users based on their roles. Give users a simple process to securely and compliantly access datasets and they will adopt it, use it, and produce better results from it.

Remember, the systems you deploy should satisfy the needs of all your personas — from data engineers and data scientists to machine learning engineers and business analysts. Finally, the results of the work performed by a user or system upstream should be made available to users and systems downstream as "data assets" that can drive business value.

Do this effectively and you can increase value from data, analytics, and AI from infrastructure savings, productivity gains, and new use cases. What can you accomplish with measures like the Databricks Lakehouse Platform? You'll:

1. **Make data more accessible** and help ensure it can be trusted.
2. **Minimize the complexity**, tools, and systems needed to perform work.
3. **Create a flywheel effect** — which will be felt across your organization and will be seen in your results — by leveraging the work of others.

INSIGHT

Make it easier for more employees to have an impact using data and AI and you'll also improve your ability to recruit and retain top talent.

databricks

6. Plan for the speed bumps to maintain business continuity

As you plan, keep in mind: the longer your data transformation takes, the more risk and cost you introduce into your organization. But it's inevitable that for a short period of time, you'll need to be "in the seam," or running both your legacy and your new systems. It's helpful to know what interim challenges this may present and plan for them now rather than as you encounter them. They can include:

- **Temporarily increasing your operational costs** while you run two sets of infrastructure
- **Raising data governance risk**, since you will have multiple copies of your data sitting in two different ecosystems
- **Adding burden to your IT workforce** due to the challenges of running multiple environments

INSIGHT

To mitigate the strain on your IT teams, consider hiring a staff augmentation service to "keep the lights on" for the legacy systems while the new systems are being implemented.

Microsoft AMD

7. Plan for a clean break from legacy platforms

Even before you're running both systems "in the seam," you'll need to know the steps and sequencing needed for shutting down your legacy platforms and finalizing the transition.

Databricks can help you transition from your legacy data storage to the Databricks Lakehouse Platform as smoothly and quickly as possible. It can help reduce cost, time-to-value, and risk for your most sensitive data and AI use cases with a simple, scalable, and unified data platform, secured and powered by Azure confidential computing on AMD EPYC™ processors with Infinity Guard featuring SEV-SNP technology.

INSIGHT

During the transition, there will be pressure to make changes to the legacy environments or to extend their life spans. Plan for these contingencies and set firm dates for when legacy systems will be retired to serve as a forcing function for teams when they onboard to the new modern data architecture.

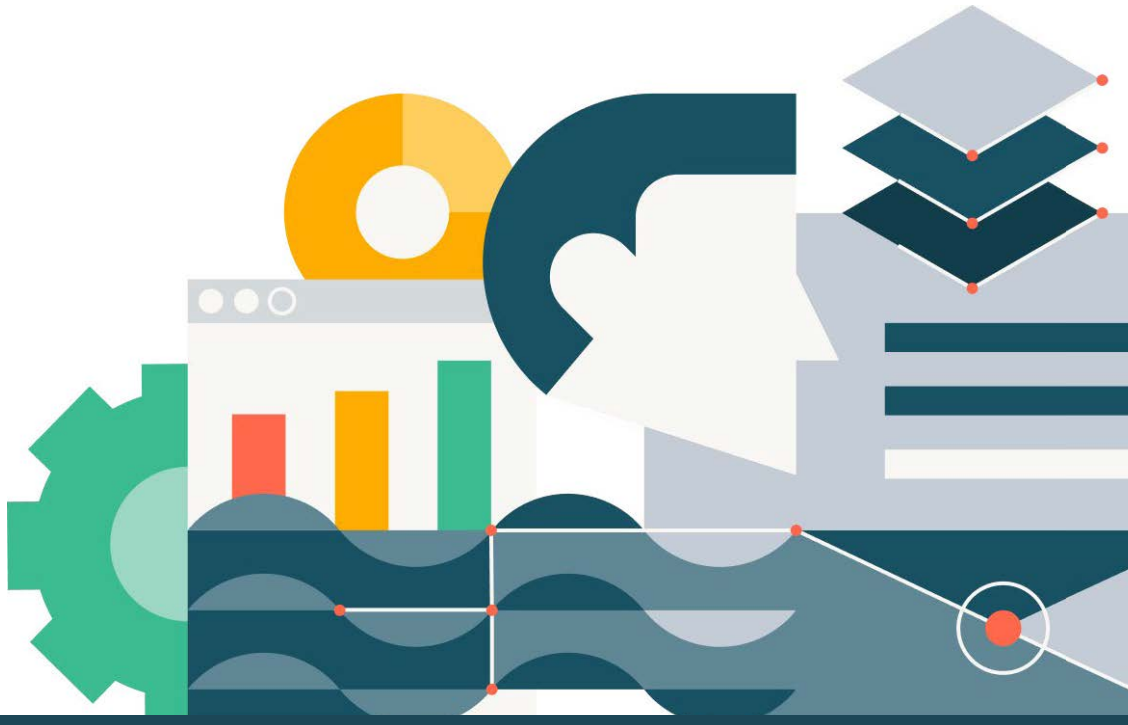
databricks

Continue your data and AI journey

Congratulations! You've taken the first step on the journey to becoming a data and AI–driven organization powered by Databricks Lakehouse Platform, Microsoft Azure, and AMD technologies.

For more resources, insights, and tools, visit the Databricks Lakehouse Platform resource hub. There you'll find resources such as videos and whitepapers.

READ MORE



Microsoft AMD