Executive Guide

Transform and Scale Your Organization With Data and AI

databricks

Contents

databricks

Process **People** **Platform**

Executive Summary

In November 2022, the release of ChatGPT, OpenAI's generative AI chatbot, signaled a significant change for organizations looking to leverage AI technologies. Overnight, ChatGPT made AI more accessible to everyone. Since then, interest in large language models (LLMs) has fundamentally changed the expectations people and businesses have in their interactions with computers and data. The latest annual McKinsey Global Survey on the state of AI confirms the explosive growth of generative AI tools. Less than a year after many of these tools debuted, one-third of survey respondents said their organizations regularly use GenAI in at least one business function.

Generative AI has the potential to disrupt every industry. Organizations therefore want to move fast in this space and accelerate innovation to differentiate themselves from the competition. Members of the C-suite everywhere are asking, "How do we accelerate our company's plan for analytics and AI? How do we start to get value from these systems as quickly as possible?" Most critically, everyone wants to bypass the hype and figure out how to build differentiated generative AI applications trained on their own data.

These changes are putting even more pressure on C-suite technology leaders already facing many challenges to deliver on their data strategy. In addition to figuring out how to deploy a modern data architecture,

leverage governed data efficiently and securely, stay compliant with an ever-increasing set of regulations, and hire the right talent, C-suite tech leaders are now under pressure to identify and execute on AI opportunities.

In our experience, technical and business leaders often underestimate the scope of changes needed to put data and AI to work. It's more than just adopting a few new IT tools, testing an AI application or moving to the cloud. To successfully lead data and AI transformation initiatives, C-suite tech executives need to develop and execute a comprehensive strategy that enables them to easily deploy a modern data architecture, unlock the full potential of all their data for analytics and AI, and future-proof their investments to provide the greatest ROI.

So what's the formula for a successful data and AI strategy? **Like so many things, it all comes down to the right process, people and platform**. Databricks has helped over 10,000 companies achieve data, analytics and AI breakthroughs. We have captured the lessons learned and summarized them in this Executive Playbook — designed to serve as a blueprint for CIOs, CDOs, CTOs and other data and AI executives to implement successful digital transformation initiatives for data, analytics and AI. This eBook takes a step-by-step approach to guide C-suite executives through critical considerations around process, people and platform. Our intention is to equip you with the knowledge to ask informed questions, make the most critical decisions early in the process, and develop a comprehensive strategy to accelerate your data and AI transformation.

databricks

Chapter 1

Process:
Establish a long–term strategy

Process

People

Platform

The most critical step to enable data, analytics and AI at scale is to develop a comprehensive and executable plan for how your organization will drive measurable business results against your corporate priorities. This strategy serves as a set of principles that every member of your organization can refer to when making decisions. To do so, first think about your goal and end state. What is your North Star?

For many organizations, it's about democratizing access to data and leveraging AI to drive innovation.

databricks

Governance is foundational to your strategy

Governance is critical to data management and democratizing access. And it's not just about data; it's about all of the assets and anything else you want to do downstream. Whether you want to run analytics, enable real-time applications or use generative AI, it all starts with data and governance. If you don't get this part right, you will fail at the rest.

It's also important to remember that governance isn't purely about security. A lot of it is about knowing the questions you should ask. Do you have the correct set of data? Is the data high quality and timely? Do the right people have access? Not everybody should have access to all of the data. Finally, how do you find the data when you need it? How do you share it securely? Data governance is also about how to make your data and the models consumable.

databricks
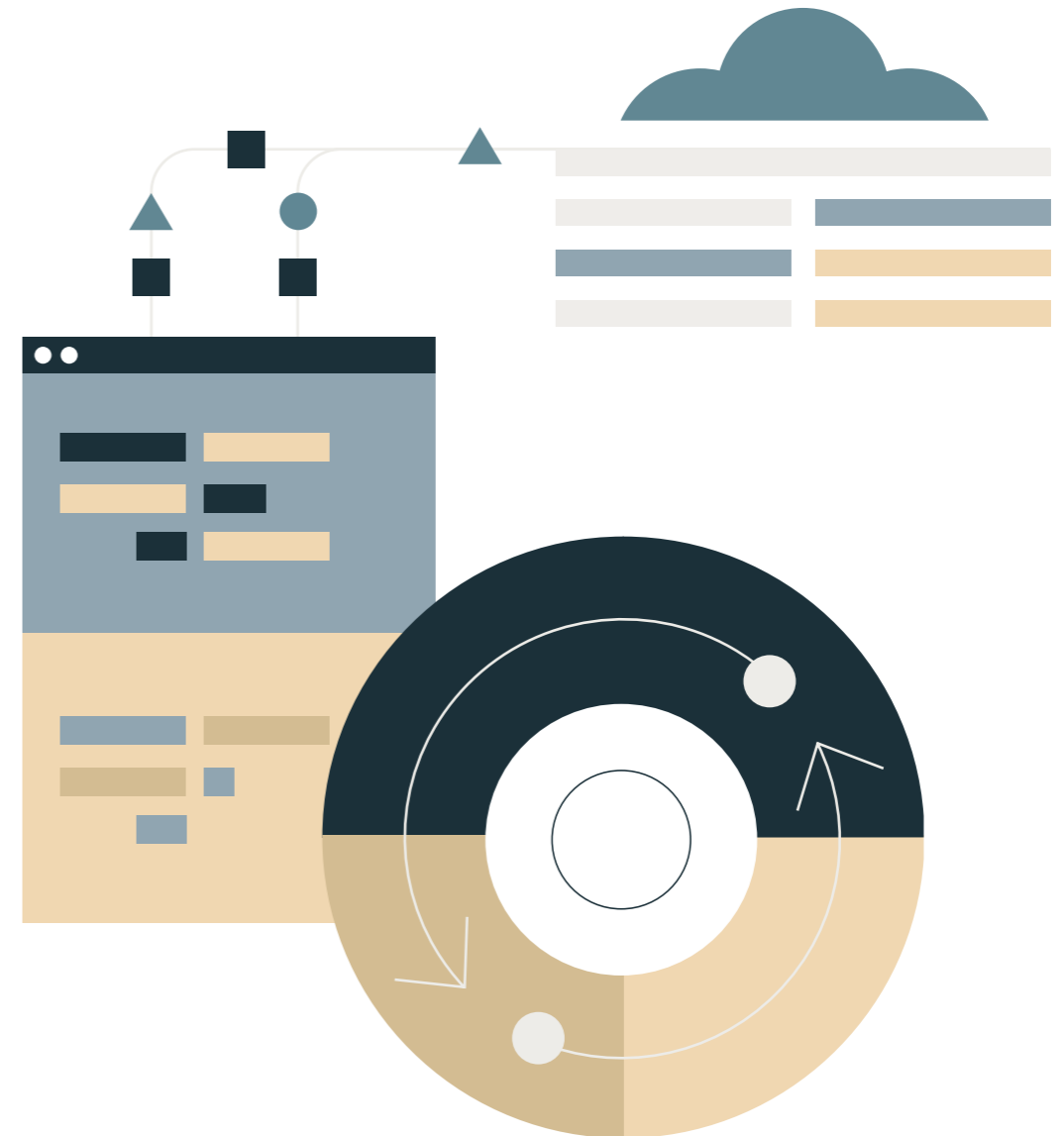
Leveraging generative AI

Generative AI and LLMs are fundamental game changers. But it's important not to chase the cool factor. Instead, start with use cases that drive business value. There are typically hundreds of use cases within an organization that could benefit from better data and AI — but not all use cases are of equal importance or feasibility. Leaders require a systematic approach to identifying, evaluating, prioritizing and implementing use cases. Many organizations start by looking at internal things around automation and human assistance.

You also need to remember to crawl, walk, run. To begin leveraging GenAI, you first need to consider things like, do you have the right dataset? What about the quality of the data? How do you get to production quality applications where output is accurate, current, aware of your enterprise context, and safe? How do you feed in your data? How do you build models? How do you understand it? You need to get your arms around those things before moving forward. And don't forget to learn from others. It's helpful to look at what peer organizations in your vertical are doing and what's worked well. The advantage of getting to things a little later on is learning what others have done so you don't make the same mistakes and can accelerate your vision. Lastly, remember that data is 100% the foundation of GenAI. You won't be successful if you don't have a good way to bring your data together and build governance around it.

databricks

How will you get there?

The second part of your strategy is determining how you will get there. As you think about your implementation path, consider what's most important to you and how you will consolidate and modernize to get there quickly. Nobody has the appetite for multiyear journeys anymore. Getting there fast is important, but you must also consider how to succeed at scale.

A key piece of your data and AI strategy involves deciding on a data platform and which components of the data ecosystem are built in-house and which components are purchased. Many engineering teams increasingly prefer to develop their own solutions in-house. This approach has some advantages — including establishing the overall product vision, prioritizing features and directly allocating the resources to build the platform. But the primary factor in making this decision should be whether a given solution offers a competitive advantage. Does a piece of software built in-house make it harder for your competitors to compete with you? If the answer is no, then it is better to focus your resources on deriving insights from your data.

Can the organization afford to wait?

If you decide the software component provides a competitive advantage and is something worth building in-house, the next question that you should ask is, "How long will it take?" There is definitely a time-to-market consideration, and the build vs. buy decision needs to also account for the impact on the business due to the anticipated delivery schedule. Keep in mind

that software development projects usually take longer, require skilled experienced engineers and cost more money than initially planned. The organization should understand the impact on the overall performance and capabilities of the daily ecosystem for any features tied to the in-house development effort. Your business partners likely do not care how the data ecosystem is built as long as it works, meets their needs, and is performant, reliable and delivered on time. Carefully weigh the trade-offs among competitive advantage, cost, features and schedule.

Don't forget about the data

The ability to make data consumable to end users and systems is perhaps the single most important feature of a data platform. Data insights, model training and model execution cannot happen reliably unless the data they depend on can be trusted and is of good quality. Focusing your efforts on curating data and creating robust and reliable pipelines provides the best chance at creating true competitive advantage. The amount of work required to properly catalog, structure, secure, ensure quality and serve up data for analysis should not be underestimated.

Finally, consider how you can maintain flexibility for the future. Picking a new platform or re-platforming is expensive and difficult. You don't want to do it multiple times.

databricks

How will you define and measure success?

Establish a clear set of metrics. Most organizations want to accelerate innovation, drive greater productivity or release new products faster so they can disrupt industries. They also want to reduce costs and minimize risk. In order to get truly game-changing results, organizations must establish a clear set of metrics to measure adoption and track the net promoter score (NPS) so that the user experience continues to improve over time. Some key metrics to keep an eye on might include things like the percentage of source systems contributing data to the ecosystem, the volume of data written to the data lake, or the number of tables defined and populated with curated data.

databricks

Chapter 2

People:
Understand your organization

To understand your organization, you first need to recognize the needs of the people in your organization. That means meeting users where they are. The key is figuring out what each group of users — whether they are data scientists, data engineers, analysts or business users — is looking to accomplish, the type of data they need and how they want to interface with that data. Understanding the people within the organization and the interface they need is an important part of change management and key to encouraging collaboration.

Once you understand the people in your organization, there are a few key things to keep in mind as you work to enable data, analytics and AI at scale:

Process

People

Platform

databricks

Balance control with autonomy

First, think about what you want to accomplish centrally and what you want to achieve in a distributed manner. Certain things should be non-negotiables from a centralized perspective. But to unite your people, you must define your core architectural vision. Next, define principles around governance and security. Once you've accomplished that, you can give people some autonomy by setting up self-service tools they can use to explore data and generate new insights on their own. For example, your supply chain might generate some random data that's valuable to the finance team. Or perhaps sales is generating data that the marketing team wants. The lakehouse architecture is great because it gives you one place to put all your data and everybody can have their own views of data within it. You can then use governance tools to publish, consume, curate and secure the data. Having a core place to put that data and a single infrastructure and platform where you can share and consume data is critical.

Empower users

As you go down the AI path, consider your enablement and talent transformation strategy. That includes both training and change management. Getting change management right is much more complex than getting the actual technology right. But it is a core part of being successful. Successful organizations have internal change agent executives who lead change management initiatives, create things like user groups and share best practices. Finally,

determine how you'll handle cultural reinforcement. The era of data and AI means there's a lot of iterative experimentation and it's moving quickly. If you don't get your people to participate, you will struggle on that journey. People want to be successful, so make sure you are giving them a path to be successful. People are worried about change because it's difficult to adapt, and they're worried they'll get left behind or their skills will become outdated. It's important to communicate to the organization where you are going, reiterate that it's a team effort, and that you are investing in resources to help them be successful.

![databricks logo]

Chapter 3

Platform:
Less is more

Process

People

Platform

As organizations have moved from large on-premises monolithic stacks to the cloud, they've unlocked an enormous amount of flexibility. But that's also added complexity, and complexity comes at a cost. Integrating multiple apps and systems is difficult. You might upgrade one solution to find another solution no longer works. Ensure the tools you invest in will truly serve you because technical debt is high. Then, map out your architecture and requirements.

A few things to consider as you do so:

databricks

Future-proof your data and AI

Deciding between proprietary solutions and open source solutions can be challenging. They both offer benefits, but all other things being equal, an open source solution provides the advantage of portability. The majority of innovation coming in the next few years will be driven by companies that are small or don't even exist today. An open source solution provides the flexibility and portability that enable an organization to change course without having to re-platform. An open source solution can help you meet today's goals and give you flexibility for tomorrow. That's why organizations prefer data platforms that let them maintain data in an open format while providing support through alternative means.

Get to the end state faster

Once you figure out the platform, how do you get to the end state faster? There are typically two options for people to take when they migrate to the cloud. The first is to lift and shift and modernize later. In almost every case, when an organization takes that approach, it's more expensive, more difficult and takes longer than expected. Ultimately, the business often has no appetite to sign up for another migration, so it's left behind. The second option is to lift, modernize and then shift. This is typically the most successful option.

Consolidate your data estate

Data estate consolidation is another important consideration. In some cases, organizations have created a separate data lake instead of modernizing it because it's easy to move data into a cloud data warehouse. The problem with that approach is that data warehouses were designed to be walled gardens. It's challenging to get your data out of them. It's also expensive. And it's not going to help any of your efforts when it comes to GenAI. It will also almost always be more costly and less performant to do those workloads in a data warehouse as opposed to using a lakehouse like Databricks. If you instead put your data in an open format, you can migrate on the back end, so users don't see anything. It will also drive costs down. And, you now have a single source of truth in an open format to power GenAI. In other words, you get the benefit of future-proofing and driving costs down without disrupting users.

Plan for production at scale

When it comes to choosing a platform, execution is important. That means determining a strategy for getting it into production, considering what it will look like at scale, and planning for change management and talent transformation. Successful organizations are the ones thinking about these things early on. It's about the process and the people and having a blueprint for how to get there. It's about making an organization successful and self-sufficient and getting there fast. But critically, it's also about laying the framework up front to ensure you've architected correctly.
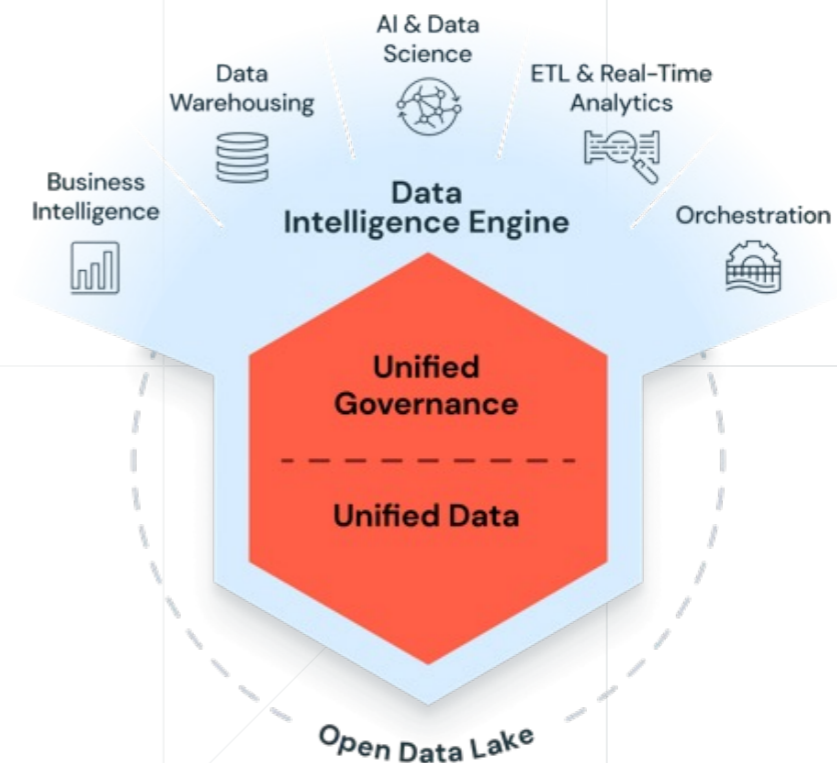
databricks

The Databricks Data Intelligence Platform

Identifying suitable data for analytics and AI initiatives poses a significant challenge for many organizations, demanding extensive curation and planning. This challenge is exacerbated by existing data platforms (or ones that are built in-house) where technical expertise is required for manual data engineering and analysis, often neglecting crucial aspects such as governance, security and privacy. Furthermore, these platforms often lack the necessary support for emerging AI applications. These issues stem from a fundamental need to comprehend how data is utilized within the organization. The emergence of GenAI opens up an opportunity for data platforms to apply intelligent techniques in addressing the challenges associated with existing data platforms.

Data intelligence, facilitated and used by AI in a data platform, enhances data management to comprehensively analyze the meaning of enterprise data. These platforms adopt the lakehouse architecture, providing a unified system for querying, governing and managing all enterprise data. By automatically analyzing data and its usage, data intelligence platforms empower organizations to develop AI applications while ensuring stringent data privacy and control.

The Databricks Data Intelligence Platform empowers organizations to leverage data and AI seamlessly. Built on lakehouse architecture, with a unified governance layer spanning data and AI, and a single query engine covering ETL, SQL, machine learning and BI, the platform brings the best of data lakes and data warehouses together. The Databricks Data Intelligence Platform combines this lakehouse foundation with

an intelligence engine that uses generative AI models to understand the semantics of your data and uses that understanding across everything in the platform. This helps reduce costs, improve productivity and enables organizations to efficiently manage, utilize and access all their data and AI. From ETL to data warehousing to generative AI, Databricks helps you simplify and accelerate your data and AI goals.

For more information about the **Databricks Data Intelligence Platform**, please visit **Databricks.com** or **contact us**.

About Databricks

Databricks is the data and AI company. More than 10,000 organizations worldwide — including Comcast, Condé Nast, Grammarly and over 50% of the Fortune 500 — rely on the Databricks Data Intelligence Platform to unify and democratize data, analytics and AI. Databricks is headquartered in San Francisco, with offices around the globe, and was founded by the original creators of Lakehouse, Apache Spark™, Delta Lake and MLflow. To learn more, follow Databricks on X, LinkedIn and Facebook.

databricks