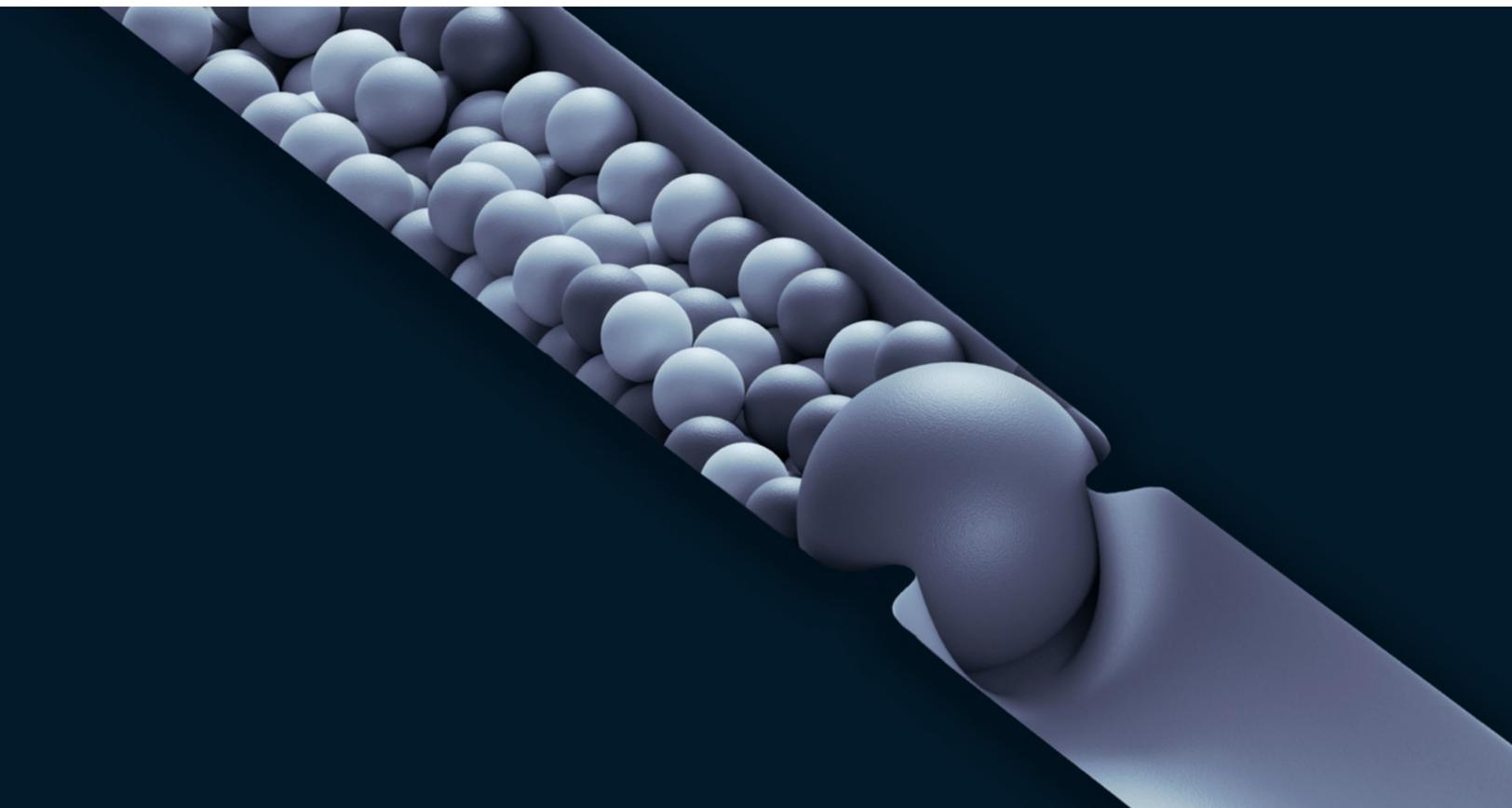# Breaking through data-architecture gridlock to scale AI

Large-scale data modernization and rapidly evolving data technologies can tie up AI transformations. Five steps give organizations a way to break through the gridlock.

by Sven Blumberg, Jorge Machado, Henning Soller, and Asin Tavakoli

© Getty Images

For today's data and technology leaders, the pressure is mounting to create a modern data architecture that fully fuels their company's digital and artificial intelligence (AI) transformations. In just two months, digital adoption vaulted five years forward amid the COVID-19 crisis. Leading AI adopters (those that attribute 20 percent or more of their organizations' earnings before interest and taxes to AI) are investing even more in AI in response to the pandemic and the ensuing acceleration of digital.

Despite the urgent call for modernization, we have seen few companies successfully making the foundational shifts necessary to drive innovation. For example, in banking, while 70 percent of financial institutions we surveyed have had a modern data-architecture road map for 18 to 24 months, almost half still have disparate data models. The majority have integrated less than 25 percent of their critical data in the target architecture. All of this can create data-quality issues, which add complexity and cost to AI development processes, and suppress the delivery of new capabilities.

Certainly, technology changes are not easy. But often, we find the culprit is not technical complexity; it's process complexity. Traditional architecture design and evaluation approaches may paralyze progress as organizations overplan and overinvest in developing road-map designs and spend months on technology assessments and vendor comparisons that often go off the rails as stakeholders debate the right path in this rapidly evolving landscape. Once organizations have a plan and are ready to implement, their efforts are often stymied as teams struggle to bring these behemoth blueprints to life and put changes into production. Amid it all, business leaders wonder what value they're getting from these efforts.

The good news is that data and technology leaders can break this gridlock by rethinking how they approach modernization efforts. This article shares five practices that leading organizations use to accelerate their modernization efforts and deliver value faster. Their work offers a proven formula for those still struggling to get their efforts on track and give their company a competitive edge.

1. Take advantage of a road-tested blueprint

Data and technology leaders no longer need to start from scratch when designing a data architecture. The past few years have seen the emergence of a reference data architecture that provides the agility to meet today's need for speed, flexibility, and innovation (Exhibit 1). It has been road-tested in hundreds of IT and data transformations across industries, and we have observed its ability to reduce costs for traditional AI use cases and enable faster time to market and better reusability of new AI initiatives.
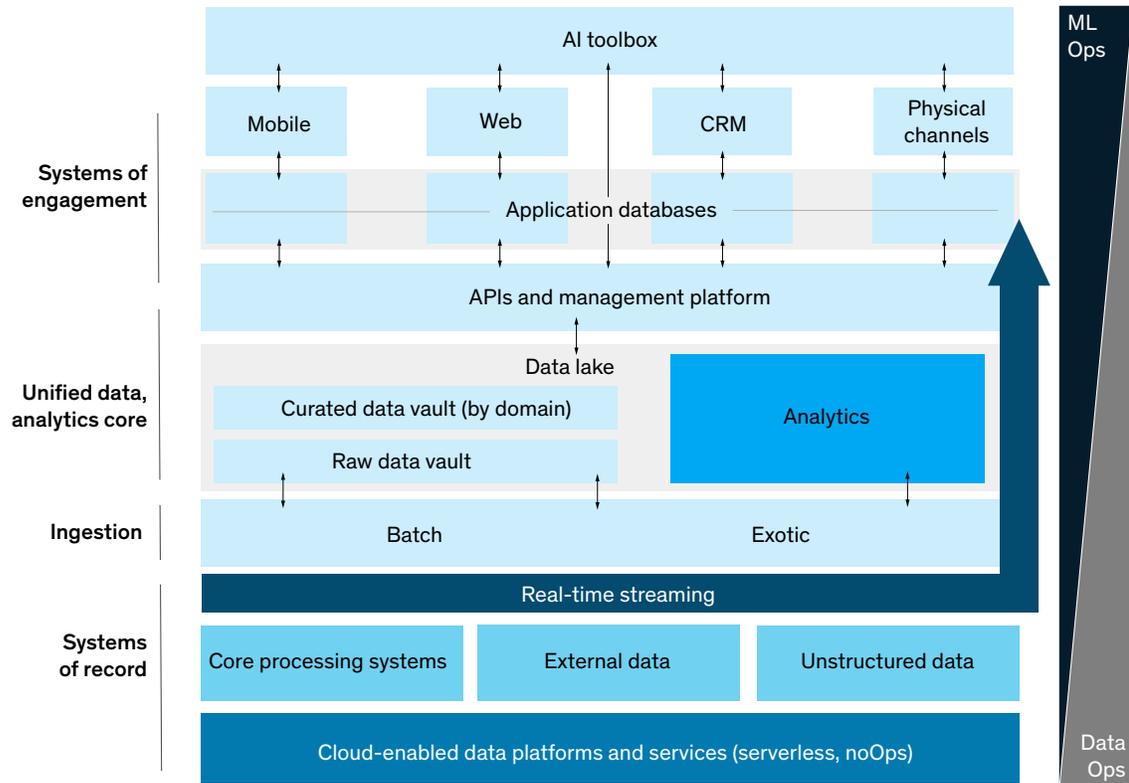
With the reference data architecture, data and technology leaders are freed from spending cycles on architecture design. Instead, leveraging this blueprint, they can iteratively build their data architecture.

Take the case of a large German bank. By using this reference data architecture as its base, the organization reduced the time required to define its data-architecture blueprint and align it with each stakeholder's needs from more than three months to only four weeks. Before adoption of the reference data architecture, business executives would become disillusioned as the CIO, CFO, risk leaders, and business executives debated architectural choices and conducted lengthy technology evaluations, even when product differences had no material impact on the bank's goals. To shift tactics, the company's CIO identified the minimal deviations required from the reference architecture and presented to all the stakeholders examples of companies across industries that had succeeded with the same approach. Executives agreed they had the setup, market positioning, and talent pool to achieve similar results, and the CIO's team quickly began building the new architecture and ingesting data.

Importantly, this isn't a one-and-done exercise. Each quarter, technology leaders should review progress, impact, funding, and alignment with strategic business plans to ensure long-term alignment and a sustainable technology build-out. One global bank implemented a new supply-based funding process that required business

Exhibit 1

A reference data architecture for AI innovation streamlines the design process.

units to reprioritize their budgets quarterly against immediate business priorities and the company's target technology road map before applying for additional funds. This new process helped the bank overcome underfunding of $250 million in the first year while gaining immediate business impact from refocused efforts.

2. Build a minimum viable product, and then scale

Organizations commonly view data-architecture transformations as "waterfall" projects. They map out every distinct phase—from building a data lake and data pipelines up to implementing data-consumption tools—and then tackle each only after

completing the previous ones. In fact, in our latest global survey on data transformation, we found that nearly three-quarters of global banks are knee-deep in such an approach.[1]

However, organizations can realize results faster by taking a use-case approach. Here, leaders build and deploy a minimum viable product that delivers the specific data components required for each desired use case (Exhibit 2). They then make adjustments as needed based on user feedback.

One leading European fashion retailer, for instance, decreased time to market of new models and reduced development costs when it focused first on the architectural components necessary for its

Exhibit 2

Each common business use case is associated with components of the data architecture.

Illustrative

Most common use cases, by component

AI tools

- Chatbots
- Marketing technology (eg, customer data platform or campaign management)
- Relationship-based pricing
- Intelligent dashboards showing spending patterns

Application programming interfaces (APIs)

- Data monetization
- Data ecosystems
- Virtual assistants
- ID proofing
- Master-data management

Data warehouse	Data lake	Data streaming
• Financial reporting (profit and loss, balance sheet) • Credit-risk reporting • Loan-application scoring	• Campaign and performance reporting • Predictive marketing • 360-degree customer view • Compliance (drawing on historical stores of multiple data types) • New use-case and model testing	• Personalization • Anti-money-laundering (AML) fraud and transaction monitoring • Real-time data ingestion

Shared ingestion layer

- Fast access and test-and-learn research and development via AI sandboxes

priority use cases. At the outset, leaders recognized that for data-science teams to personalize offerings effectively across multiple online and mobile channels, including social channels, they would need fast access to data. Previously, data scientists had to request data extracts from IT, and data were often outdated when received.

The retailer's focus on the architecture its use cases required enabled development of a highly automated, cloud-based sandbox environment that provides fast access to data extracted from a shared, company-wide ingestion layer; an efficient manner to spin up analytics and AI sandboxes

as needed; and a process to shut them down when they aren't needed. Whereas physical and virtual environments could once run up IT bills for months and years, such environments can now be accessed on the cloud for less than 30 minutes—the average amount of time that they're actually needed—generating substantial cost savings.

Once organizations finish building the components for each use case, they can then scale and expand capabilities horizontally to support other use cases across the entire domain. In the case of the retailer, as

new personalized offerings become ready for deployment, the organization moves the selected data features into curated, high-quality data environments for production access.

3. Prepare your business for change

Legitimate business concerns over the impact any changes might have on traditional workloads can slow modernization efforts to a crawl. Companies often spend significant time comparing the risks, trade-offs, and business outputs of new and legacy technologies to prove out the new technology.

However, we find that legacy solutions cannot match the business performance, cost savings, or reduced risks of modern technology, such as data lakes. Additionally, legacy solutions won't enable businesses to achieve their full potential, such as the 70 percent cost reduction and greater flexibility in data use that numerous banks have achieved from adopting a data-lake infrastructure for their ingestion layer.

As a result, rather than engaging in detailed evaluations against legacy solutions, data and technology leaders better serve their organization by educating business leaders on the need to let go of legacy technologies. One telecom provider, for example, set up mandatory technology courses for its top 300 business managers to increase their data and technology literacy and facilitate decision making. As part of the training, the data leadership team (including engineers, scientists, and practitioners) shared the organization's new data operating model, recent technology advances, and target data architecture to help provide context for the work.

In addition to educating business leaders, organizations should refocus efforts from their legacy stack to building new capabilities, particularly in the infrastructure-as-a-service space. A chemical company in Eastern Europe, for instance, created a data-as-a-service environment, offloading large parts of its existing enterprise resource planning and data-warehouse setup to a new cloud-based data lake and provisioning the underlying data through standardized application programming interfaces (APIs). This approach reduced time to market and made it easier to use fast-paced analytical modeling, enabling new customer-360 and master-data-management use cases, while reducing the complexity of the overall environment.

4. Build an agile data-engineering organization

In our experience, successful modernization efforts have an integrated team and an engineering culture centered around data to accelerate implementation of new architectural components. Achieving this requires the right structural and cultural elements.

From an organizational perspective, we see a push toward reorienting the data organization toward a product and platform model, with two types of teams:

— Data platform teams, consisting of data engineers, data architects, data stewards, and data modelers, build and operate the architecture. They focus on ingesting and modeling data, automating pipelines, and building standard APIs for consumption, while ensuring high availability of data, such as customer data.

— Data product teams, consisting mostly of data scientists, translators, and business analysts, focus on the use of data in business-driven AI use cases such as campaign management. (To see how this structure enables efficiency across even the larger, more complex organizations, see sidebar, "Sharing data across subsidiaries.")

The cultural elements are aimed at improving talent recruiting and management to ensure engineers are learning and growing. A Western European bank is cultivating a learning culture through a wide range of efforts:

— *Providing engineers with clearly documented career paths.* This includes establishing formal job levels for engineers based on their productivity, with promotion rounds based on qualitative feedback, their contributions to open-source communities, their management skills, and their knowledge, all assessed against

Sharing data across subsidiaries

Across industries, regulators and companies' risk, compliance, supply chain, and finance departments are increasingly asking for granular data access covering the headquarters and subsidiaries. On the regulatory side, for example, companies exporting products that can be used for both civilian and military applications must provide regulators full transparency across the value chain. On the operational side, such transparency can help provide more advanced insight into global supply chains and operations and improve productivity, reducing the resources needed to build and manage an end-to-end data architecture in every country.

In response, organizations are moving toward defining data-architecture strategies that can transfer learnings from headquarters to subsidiaries or vice versa. Companies that do this well, such as Amazon, Google, and Microsoft, harmonize their business and technology delivery models. This entails setting up a global team with a clear product owner, who owns the global data model, and dedicated data architects and engineers, who create a shared data vault containing the granular transaction data of the subsidiaries. Local engineers within the subsidiaries then make any customizations they need while remaining aligned with global teams.

By taking this approach, a French bank drastically improved the quality of its anti-money-laundering and know-your-customer reporting while lowering the cost of the data architecture for subsidiaries by 30 percent. These positive results have laid the foundation for groupwide scaling of another data lake to support other use cases, such as calculating risk.

a structured maturity grid. The bank also revised its compensation structure to ensure that engineers at the highest job levels receive compensation comparable to that of senior managers in IT, data, and the business.

— *Adopting a pragmatic approach to assessing expertise levels.* Research indicates that expert engineers are eight times more productive than novices, so the success of modernization efforts depends on effective recruitment, management, and organization of talent. To provide a consistent measurement for recruiting, upskilling, and advancement, the bank used the well-known Dreyfus model for skill acquisition to identify five aptitude levels from novice to master, rate observable behavior through key indicators, and develop individual training plans based on the feedback.

— *Establishing a culture of continuous technology learning.* Continuous learning requires the sharing of expertise through formal and informal forums, peer reviews, and freedom to pursue online training courses, certifications,

and virtual conferences. To support this, bank leaders have instituted an agile performance-management model that emphasizes both knowledge and expertise. At other organizations, the performance measurement of top executives and team members includes their industry contributions; their success metrics might include, for example, the number of keynote presentations they deliver throughout the year.

— *Emphasizing engineering skills and achievements.* To emphasize technical skills, the bank encourages everyone in IT, including managers, to write code. This creates a spirit of craftmanship around data and engineering and generates excitement about innovation.

5. Automate deployment using DataOps

Changing the data architecture and associated data models and pipelines is a cumbersome activity. A big chunk of engineering time is spent on reconstructing extract, transform, and load

(ETL) processes after architectural changes have been made or reconfiguring AI models to meet new data structures. A method that aims to change this is DataOps, which applies a DevOps approach to data, just as MLOps applies a DevOps approach to AI. Like DevOps, DataOps is structured into continuous integration and deployment phases with a focus on eliminating "low-value" and automatable activities from engineers' to-do lists and spanning the delivery life cycle across development, testing, deployment, and monitoring stages. Instead of assessing code quality or managing test data or data quality, engineers should focus their time on code building. A structured and automated pipeline, leveraging synthetic data and machine learning for data quality, can bring code and accompanying ETL and data-model changes into production much faster.

One large pharmaceutical company is working to bring biometric insights to its front line more quickly using DataOps. It has defined automated ways to test new biometric analytics models against standards and developed a code library to optimize code reuse. It is currently defining an easier way to deploy models in production to reduce time lags between model development and use. Once completed, this will reduce the typical time required to deploy models and apply results, such as identifying the right mixtures, from weeks to hours.

———————

Today, most data technologies are readily available in the cloud, making adoption a commodity. As a result, the difference between leaders and laggards in the data space will depend on their ability to evolve their data architecture at a brisk pace to harness the wealth of data collected over decades and new data streaming in. Organizations that can't move as quickly risk derailing their digital and AI transformations. The five practices we have outlined, along with a positive vision and a compelling story for change, can enable organizations to move at the necessary speed, building momentum and value along the way.

Sven Blumberg is a senior partner in McKinsey's Istanbul office, **Jorge Machado** is a partner in the New York office, **Henning Soller** is a partner in the Frankfurt office, and **Asin Tavakoli** is a partner in the Düsseldorf office.