Enterprise Strategy Group™ by TechTarget

WHITE PAPER

Data Readiness for SAP Applications: Capitalizing on AI and GenAI to Drive Enterprises Forward

Utilizing Fivetran, Databricks, Amazon Web Services, and Capgemini to Revolutionize Data for the Enterprise

By Stephen Catanzano, Senior Analyst
Enterprise Strategy Group

June 2024

Contents

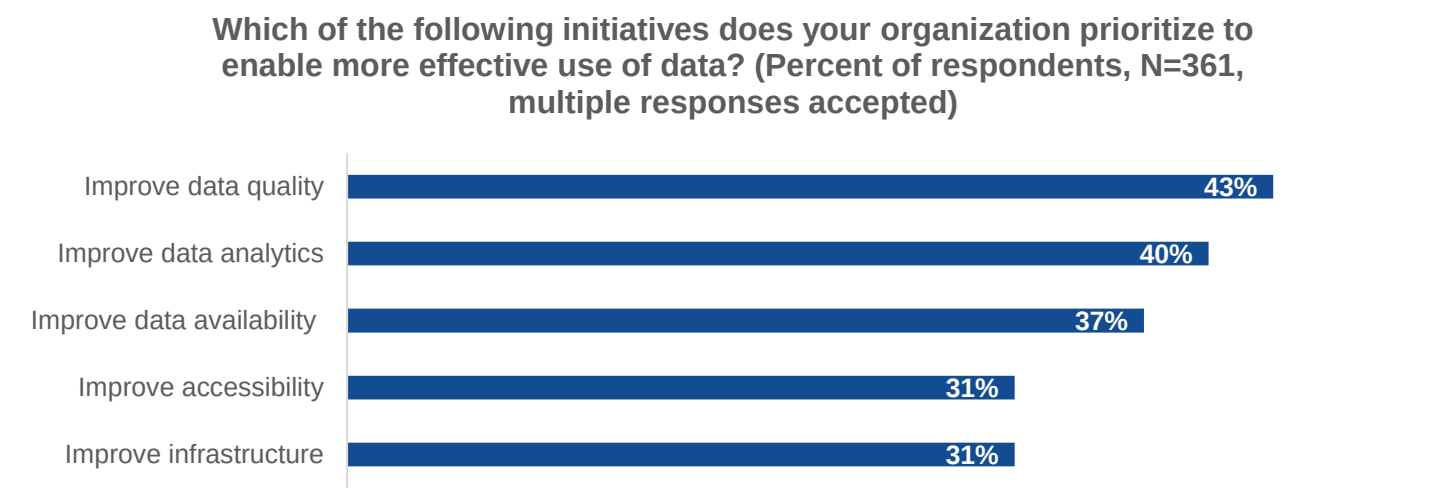
Abstract

The advent of AI and generative AI (GenAI) represents a transformative leap in the capabilities of organizations to leverage data to create unprecedented innovation and efficiencies. Data readiness encompasses an organization's processes and strategies to ensure that their data assets are accurate, accessible, and actionable for AI-driven applications. Data-driven organizations find both challenges and a symbiotic relationship between their data and AI and GenAI models, which all demand extremely well-prepared data environments. Organizations implementing AI will see a transformative impact on operational efficiency, customer experiences, and innovation when a robust data readiness strategy underpins data used in AI models. This white paper focuses on the power of SAP data as a core source—with data readiness applied—to not only power SAP applications but also be a foundation for AI and GenAI applications within the organization. This can fundamentally shift how SAP data can be valued, managed, and utilized to transform an organization. Fivetran, Databricks, Amazon Web Services (AWS), and Capgemini provide guidance on maximizing the value of SAP data to build unique AI-powered solutions.

Data Empowerment in Organizations Continues to Climb

Organizations are rapidly attempting to democratize the use of data across their organization and empower their workforce with accurate and trusted data to drive the business forward. As seen in Figure 1, TechTarget's Enterprise Strategy Group asked research participants what their data usage priorities were focused on to enable users to be more effective and found that 43% prioritized improving data quality as a top priority. This most commonly cited response was followed by improving data analytics (cited by 40%), improving data accessibility (40%), improving accessibility of data (31%), and improving infrastructure to manage data (31%).[1] Enterprise Strategy Group research also found that 84% of organizations saw an increase in end-user access to their data in the past year and that 53% of line-of-business stakeholders only somewhat trust the accuracy of the data given to end users for decision-making.[2]

Figure 1. Top Priorities for Enabling End Users to Use Data More Effectively



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

[1] Source: Enterprise Strategy Group Complete Survey Results, *The State of DataOps: Unleashing the Power of Data*, December 2023.
[2] Ibid.

It's clear that organizations understand that leveraging data from across the organization to empower employees is crucial for long-term success. The goal of being data-driven is clear, but achieving the goal involves having excellence in data management, using AI and Gen AI as accelerators, and building a foundation with trusted data. As organizations democratize, trust—including accuracy, timeliness, quality, and governance—becomes part of the foundation of data readiness.

Within data derived from SAP applications, we find core operational data in finance, accounting, supply chain, human resources, sales, marketing, and more. This current and historical data is truly valuable as part of the data foundation for AI and GenAI models and the building of a trusted data source.

The Motivation for Modernizing

Organizations are under pressure to prepare for the future and build resilient, data-driven companies. Many organizations have shifted from the idea that data is part of their business to it now *being* the business. This shift is driven by market pressures from the board level, consumers, investors, suppliers, and senior-level managers, as the reality of AI is here and no longer a future consideration. However, for effective AI applications, data must be relevant, high-quality, and timely. This is one of the core challenges organizations have to meet in order to achieve the desired outcomes. The good news is that the goals are clear, with the challenges focused on the data, tools, and processes needed to achieve them. Some of the motivations to implement AI across organizations include:

- **Accelerating growth.** Intelligent data drives faster decision-making, resulting in transformative growth.
- **Streamlining operations.** Innovation happens faster when processes are streamlined.
- **Improved customer engagement.** Higher customer satisfaction, increased retention, higher average sales, and loyalty all help to keep businesses afloat in a competitive environment.
- **Competitive advantage.** Staying ahead of the competition is achieved with differentiation as well as innovation.

Market Insight

79% of organizations say they must use AI in business- and mission-critical processes to compete better.

AI helps to achieve these goals. As organizations strive to modernize their applications, they are looking to integrate AI in order to position themselves for the future. In a recent Enterprise Strategy Group research survey, we asked organizations about their plans for AI, and 79% say they must use AI in business- and mission-critical processes to compete better.[3]

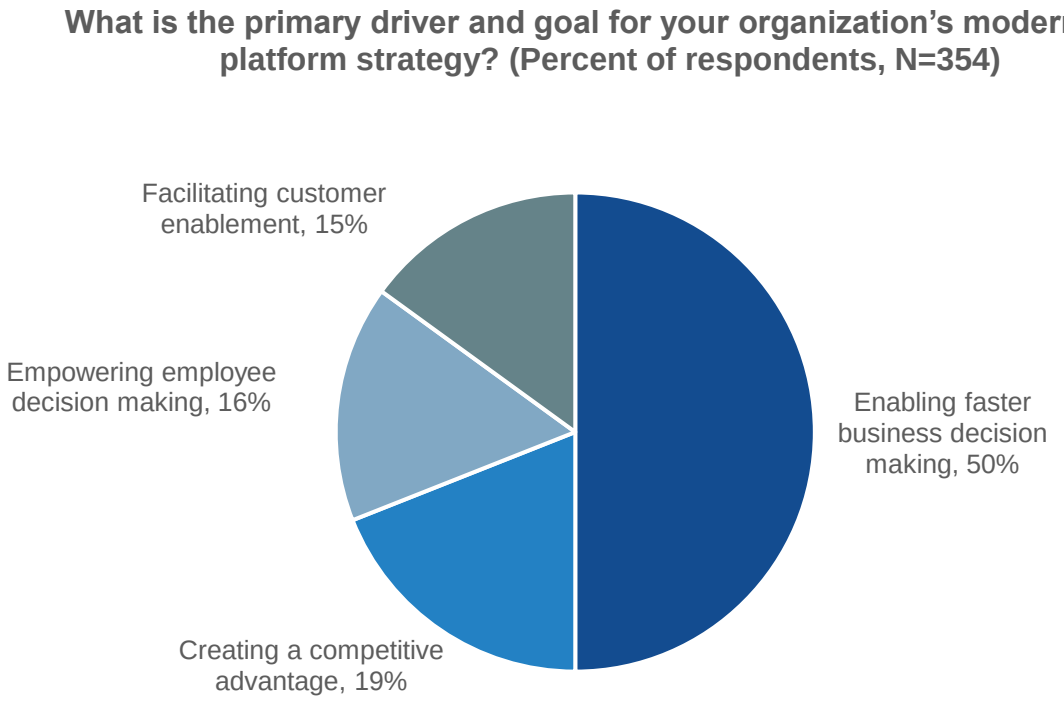
With data at the core of remaining competitive and the future success of an organization, the data platform becomes increasingly important. This is an architecture of how data is integrated into the organization; transformed via extract, transform, and load (ETL) pipelines; stored (via data lakes, data warehouses, databases, etc.); used by machine learning (ML), AI, and GenAI models and processes; and, ultimately, visualized by applications. SAP applications are at the core of many organizations, both creating new data and visualizing data to empower employees and the business.

Modern data platforms are helping organizations think about their data from an end-to-end perspective. They are integral to building the trusted data source used as the foundation for AI and the intelligence fuel in every organization. As shown in Figure 2, Enterprise Strategy Group asked research participants what the primary driver and goal was for their modern data platform strategy and found that 50% were focused on enabling faster business decision-making, 19% on creating a competitive advantage, 16% on empowering employee decision-making, and 15% on facilitaing customer enablement.[4]

[3] Ibid.

[4] Source: Enterprise Strategy Group Research Report, *Data Platforms: The Path to Achieving Data-driven Empowerment*, July 2023.

Figure 2. Organizations Understand the Value Data Brings



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Utilizing SAP-derived Data to Power Unique AI Solutions

At the core of every SAP application—including SAP S/4 HANA, SAP CRM, SAP ERP Central Component (ECC), SAP Business Warehouse (BW), and SAP SuccessFactors—lies incredibly valuable insights that help power an organization. Unlocking and enhancing these insights to use within the organization can quickly lead to a new level of data empowerment.

As organizations modernize their SAP applications by moving them to the cloud to take advantage of scalability and efficiencies as well as leverage rapidly expanding AI capabilities, they should consider utilizing their SAP data as part of their foundational data for AI applications, helping to democratize this valuable data resource.

The vast and insightful data from SAP applications can be utilized in AI and Gen AI models in various ways to enhance their business operations, improve decision-making, and drive innovation.

SAP-derived data can help in AI and GenAI applications developed by an organization or utilized in third-party solutions. Following are some examples of how SAP-derived data can be used to build or enhance large language models (LLMs), extract the value of the data, and empower the organization. This operational data is extremely relevant, especially when building domain-specific models that leverage proprietary data and truly create competitive advantage and innovation. SAP-derived data provides:

- **Predictive analytics.** Organizations can leverage SAP data to integrate with AI-powered analytic capabilities. These functionalities can analyze historical data and predict future trends, enabling organizations to make more informed decisions in areas such as sales forecasting, demand planning, and inventory management.

- **Intelligent automation.** SAP data can enable AI-powered automation to help streamline repetitive tasks and processes. Robotic process automation (RPA) can be used to automate data entry, invoice processing, and other routine activities, freeing up employees to focus on more strategic tasks.
- **Personalized recommendations.** SAP data is a source of customer data that can be used to build AI algorithms to analyze customer data and deliver personalized recommendations and offers. This may include browsing data and purchasing history, enhancing the customer shopping experience, and increasing sales.
- **Chatbots and virtual assistants.** The data from SAP applications can be used to build a knowledge base to create AI-powered chatbots and virtual assistants for customer support, answer inquiries, and assist users in navigating information. These AI-driven interfaces can improve user engagement and satisfaction while reducing the burden on support staff.
- **Anomaly detection.** AI algorithms can detect anomalies and patterns in data. For example, financial data from SAP can be used to help customers identify potential issues or opportunities and patterns that might prevent system downtimes.

Achieving Data Readiness

When modernizing by moving SAP applications to the cloud or enhancing existing SAP applications in the cloud, organizations should carefully plan and address several key aspects, including data integration, data lakes infrastructure, and system integration services. In this new age of AI, the goal of modernizing should be to not only give SAP applications the benefits of the scale and efficiencies of the cloud, but also to have an environment where SAP-derived data is efficiently utilized for AI and GenAI applications as part of the organization's data foundation.

The data from an organization's SAP applications, when made accessible, can be utilized to train LLMs in AI. This represents a tremendous opportunity, especially during the migration of SAP to a cloud such as AWS. Building domain-specific LLMs utilizing a trusted data source with vast amounts of current and historical data, trends, and insights—like SAP-derived data—can be transformative. As with data being used in AI modeling, steps still need to be taken to ensure accuracy, quality, and governance for data readiness.

To help organizations utilize their SAP data to transform their business, Fivetran, Databricks, AWS, and Capgemini offer solutions that organizations can utilize in their journey toward building competitive AI-powered organizations.

Fivetran: Data Integration

Fivetran is the global leader in data movement. The Fivetran platform reliably and securely centralizes data from over 500 SaaS applications and databases, including SAP, into and across cloud data platforms like Databricks and AWS. Fivetran's fully managed platform has automated the most time-consuming parts of the extract, load, and transform (ELT) process. Data pipelines are prone to breaking, making it difficult for organizations to move data and integrate the data sources with their destinations. With Fivetran's industry-leading 99.9% uptime, organizations can be assured that their pipelines are always running on schedule and that they will have the freshest data.

According to an Enterprise Strategy Group research study, 32% of respondents said that their organization's most important expected outcome from implementing data integration tools or services was increased speed and performance, 32% reported increased data insights, 18% cited increased data end-user satisfaction, 9% expected reduced costs, and 9% said reduced risk.[5]

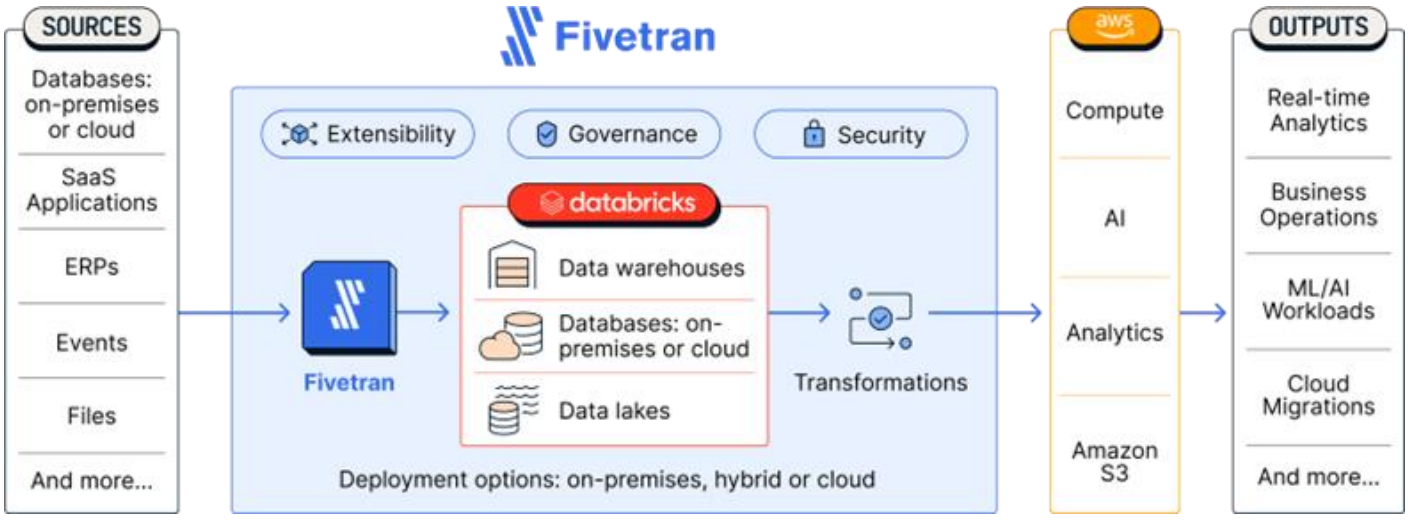
[5] Ibid.

Fivetran normalizes, standardizes, and securely centralizes SAP data with low latency and low footprint. As shown in Figure 3, Fivetran moves and transforms data from hundreds of sources to various destinations, including data lakes, data warehouses, and databases. It offers multiple SAP replication options—all of which provide continuous, real-time change data capture from operational database-based systems, including S/4HANA and SAP ECC instances running on Oracle, Db2, SQL Server, and HANA DB. In addition, the platform offers various connection options for ELT database replication, such as safelisting Fivetran's IP address, Secure Socket Shell (SSH) tunnel, reverse SSH tunnel, or private links. Fivetran can replicate data from SAP ERP instances running on Oracle, Db2, SQL Server, and the SAP HANA database, as well as from SAP S/4HANA. Fivetran's scalability enables organizations to handle the most extensive data loads with core features to maintain extensibility, governance, and security, all of which are critical elements of data readiness to prepare data for use in AI and GenAI applications.

Top Expected Outcomes From Implementing Data Ingestion Tools

- Increased speed and performance (32%)
- Increased data insights (32%)
- Increased data end-user satisfaction (18%)
- Reduced costs (9%)
- Reduced risk (9%)

Figure 3. Fivetran Data Movement Model

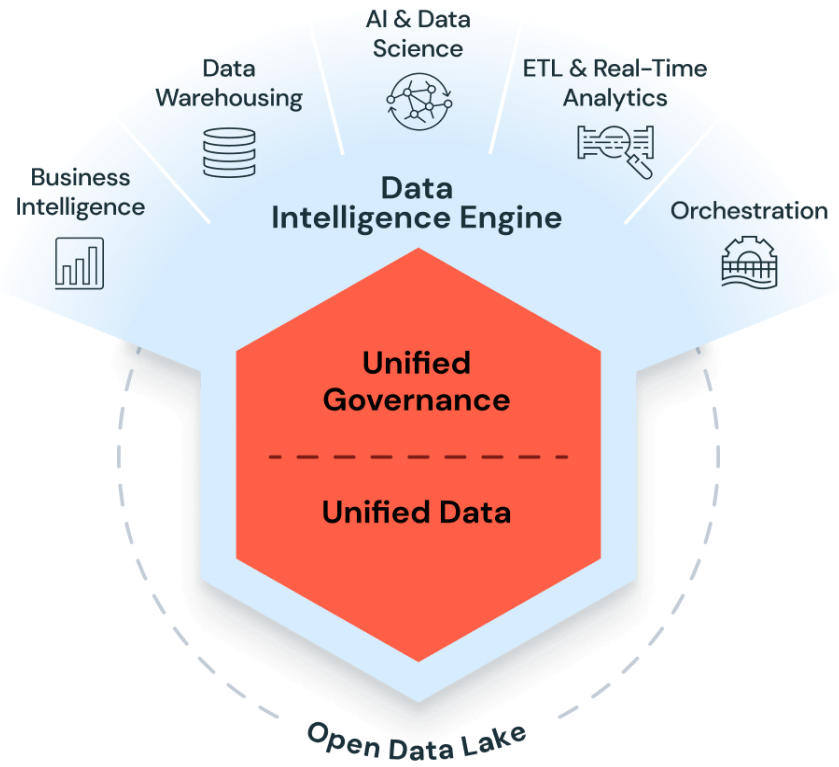


Source: Fivetran

Databricks: Unifying SAP Data with Non-SAP Data at Scale

As shown in Figure 4, the Databricks Data Intelligence Platform enables organizations to unify data and governance. It revolutionizes data management by employing AI models to deeply understand enterprise data semantics. The foundation is built on a data lakehouse, which is a unified system to query and manage all data across the organization. It analyzes both the data contents and metadata, as well as how they are used (queries, reports, lineage, etc.). This gives organizations a deep understanding of their data and how it can be applied across an organization, making it the ideal platform for SAP-derived data for SAP applications, as well as AI and GenAI applications.

Figure 4. Databricks Data Intelligence Platform



Source: Databricks.

Some of the key features of the Databricks Data Intelligence Platform include:

- **Ability to infuse GenAI.** Using the Data Intelligence Engine, the Data Intelligence Platform infuses GenAI into the workflow to understand the uniqueness of an organization's data. It observes how data is used in existing workloads to learn the organization's terms and offers a tailored natural language interface to all users, from nonexperts to data engineers. DatabricksIQ is the data intelligence engine powering the Databricks Platform. It is a compound AI system that combines the use of AI models, retrieval, ranking, and personalization systems to understand the semantics of an organization's data and usage patterns. DatabricksIQ does not have an end-user UI, but it enables existing product experiences to be more relevant and accurate, such as Databricks Assistant, AI-generated comments, and intelligent search. DatabricksIQ also enables autonomous data management and administration by automatically configuring and tuning the Databricks Platform. This improves performance, reduces manual administration, and lowers overall total cost of ownership.
- **Unified data on a cost-effective platform.** Organizations can optimize costs and improve ROI with automation and optimizations provided within the Data Intelligence Platform, which harmonizes data, including SAP-derived data, to power AI and Gen AI solutions, business intelligence, and more. Semantic cataloging and discovery are also used to understand each organization's data model, metrics, and KPIs.
- **Unity Catalog.** Databricks Unity Catalog offers a unified governance layer for data and AI within the Databricks Data Intelligence Platform. The Unity Catalog provides organizations with the following:
  - **Unified view of the data estate.** The ability to centrally catalog all data, analytics, and AI assets, including files, tables, ML models, and ML features, across clouds, regions, and platforms, makes it easy for data teams to discover, access, and analyze information.

  - **Single permission model for data and AI.** A single permission model for all data, analytics, and AI assets, no matter where they live, across clouds, regions, or platforms, provides consistent and simplified access policy management.
  - **AI-powered monitoring and observability.** Unity Catalog enables organizations to harness the power of AI to automate monitoring, diagnose errors, and maintain the quality of data and ML models. Organizations also benefit from proactive alerts that quickly detect personally identifiable information, track model drift, and resolve issues within their data and AI pipelines to ensure accuracy and integrity.
- **Enhanced governance and privacy.** The platform automatically detects, classifies, and prevents misuse of sensitive data while simplifying management by using natural language.

For SAP applications and data, the Databricks Data Intelligence Platform unifies and enhances the use of data to develop and empower AI and GenAI applications. These applications deliver the insights organizations need to build data-driven, competitive businesses.

AWS: Innovative Cloud Infrastructure

AWS is a comprehensive and widely adopted cloud platform that provides the infrastructure for organizations and technology partners like Fivetran and Databricks, including computing power, storage options, and networking capabilities. AWS can help organizations with their AI initiative-related data readiness in the following areas:

- **Compute:** Services like Amazon Elastic Compute Cloud (EC2) for scalable virtual servers, AWS Lambda for serverless computing, and Amazon Elastic Container Service (ECS) for container management.
- **Storage:** Solutions like Amazon Simple Storage Service (S3) for scalable object storage, Amazon Elastic Block Store (EBS) for block storage, and Amazon Glacier for archival storage.
- **Database:** Managed databases like Amazon Relational Database Service (RDS), Amazon Neptune (Graph Database), and Amazon DynamoDB (NoSQL database).
- **Networking:** Tools like Amazon Virtual Private Cloud (VPC) for isolated network environments, AWS Direct Connect for dedicated network connections, and Amazon Route 53 for domain name system web services.
- **Machine learning and AI:** Services such as Amazon SageMaker for building, training, and deploying machine learning models and Amazon Bedrock for access to LLMs and FMs.
- **Security and identity:** Features like AWS Identity and Access Management (IAM) for managing user permissions, AWS Key Management Service (KMS) for encryption keys, and AWS Shield for DDoS protection.
- **Developer tools:** Resources like AWS CodeCommit for source control, AWS CodeBuild for continuous integration, and AWS CodePipeline for continuous delivery.
- **Global infrastructure:** AWS has a global network of data centers organized into regions and availability zones (regions are geographic areas that contain multiple isolated locations known as availability zones). This design ensures high availability, fault tolerance, and low latency for applications.
- **Scalability and flexibility:** AWS enables businesses to scale their infrastructure up or down based on demand. Users can easily add or remove resources as needed, enabling cost-effective and efficient service use. The pay-as-you-go pricing model ensures that users only pay for the resources they consume.
- **Security and compliance:** AWS offers robust security features to protect data and applications. These include encryption, network firewalls, identity management, and compliance with global regulatory standards like GDPR, HIPAA, and SOC.
- **Innovation and ecosystem:** AWS continually innovates and introduces new services and features. It has a large ecosystem of partners, including independent software vendors (ISVs) and systems integrators (SIs), providing additional solutions and integrations.

The architecture empowers users of all levels to support their AI use cases.

Capgemini: Digital Transformation Expertise

Capgemini has four decades of experience with SAP solutions, serving 1,800 clients worldwide. It is a world leader in SAP certifications, an SAP Global Strategic Services Partner, and an SAP Global Platinum Reseller Partner. Additionally, Capgemini provides information technology, consulting, and outsourcing services. The company also offers digital services, technology solutions, cloud services, AI solutions, and cybersecurity and engineering services.

In addition to its SAP practice, Capgemini's Data and Artificial Intelligence practice enables organizations to work with their data and apply the transformative power of data, AI, and GenAI. Organizations becoming more data-powered adapt to the changing world faster and can achieve a sustainable competitive advantage. With the world becoming increasingly intelligent and entire value chains being transformed, data and AI are essential for the understanding and insight organizations need. Capgemini can help organizations extract value from their data.

Capgemini has the experience to help with SAP migrations, SAP strategy, operationalizing SAP, and building a plan for data readiness using both SAP and non-SAP data to create the foundation of data needed to create AI and GenAI solutions that can transform organizations and create competitive advantages and innovations. Capgemini has the experience to help drive the outcomes organizations desire.

Conclusion

Organizations should recognize the tremendous value of their SAP data as foundational to ensure its quality, consistency, and accessibility for AI-driven initiatives as part of a data readiness strategy. AI-driven applications and services can be developed that leverage SAP data to drive innovation further, improve efficiency, build competitive advantages, and create new business opportunities.

To address these requirements for data readiness, we have outlined some of Fivetran, Databricks, Amazon Web Services, and Capgemini's core capabilities. These companies have solutions designed to integrate SAP data into foundational data and help organizations accelerate the extraction of its value for business.

We highly recommend that organizations engage with these partners to have deeper discussions on how they can become data-driven organizations and leverage their data, AI, and GenAI initiatives to position themselves for the new world of AI-driven businesses.

About Enterprise Strategy Group
TechTarget's Enterprise Strategy Group provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

contact@esg-global.com
www.esg-global.com