



# Imperatives for affordable AI experimentation

An AI leader's guide to harnessing a  
reimagined modern data lake

available in  
 aws marketplace

# Table of contents

Introduction	3
Data lakes present new opportunities	4
Impact of data lakes on AI innovation	5
Common challenges for data lake adoption	6
Fivetran helps Tinito build a scalable data lake for AI-driven marketing	7
Overcoming legacy data infrastructure limitations	8
Ensuring data quality and pipeline reliability	9
Simplifying software purchasing and implementation	10
Conclusion	11

# Introduction

Significant advancements in data lake technology and the rise in AI use cases are transforming how data lakes are viewed. This is prompting organizations to reconsider corporate data warehouses. The original concept of a data warehouse, with components such as storage, table structure, compute/query engine, and catalog of tables, can be rigid and expensive. In a traditional data stack, data is duplicated to multiple warehouse destinations to serve the needs of different departments and use cases. This approach increases complexity and creates additional compliance risks while increasing the total cost of ownership (TCO).

The rise in popularity of data lakes stems from the need of organizations to centralize and access high volumes of structured and unstructured data to deliver evolving AI use cases. Modern data lakes simplify the catalog component with API-based data catalogs and open table formats, making data more accessible. Leading companies embrace these shifts to become innovators, providing data lakes that deliver advanced analytics and AI/ML use cases without manual intervention. Fivetran is one such innovator, and with the Fivetran Managed Data Lake Service, customers benefit from maintenance-free, high-quality data pipelines.

This eBook explores how modern data lakes support an organization's evolving infrastructure without the complexity seen in the past.

90%

of IT leaders aim to consolidate analytics into a single location<sup>1</sup>

85%

of organizations are using their data lakes to support AI model development<sup>1</sup>

56%

of organizations report saving on analytic costs by moving to a data lake<sup>2</sup>

1. [New "State of the Data Lakehouse in the AI Era" Report](#)
2. [whitepaper-2024-state-of-the-data-lakehouse\\_report.pdf](#)

# Data lakes present new opportunities

Why are organizations increasingly adopting data lakes over traditional data warehouses? Data lakes are cost effective, scalable, and flexible. They can handle all kinds of data—structured, semi-structured, and unstructured. This makes them perfect for AI and machine learning projects.



AI requires vast amounts of high-quality data, but getting that data into the right format at scale is a major challenge,"

— [George Fraser, CEO of Fivetran](#).

"Our Managed Data Lake Service for Cloud Storage automates the entire process — moving, organizing, and optimizing data in open table formats — so businesses can focus on leveraging insights and driving innovation."

However, some organizations remain hesitant to transition to data lakes due to legacy perceptions and operational concerns. Many technology leaders still associate data lakes with outdated architectures that lacked governance and performance optimization. This led some to referring to a data lake as a data swamp, making it difficult to locate or query records. Some leaders also believe that their current data warehouses are sufficient for business intelligence needs. Real-time data requirements and the perceived complexity of managing data lakes—especially around ingestion and schema evolution—continue to be barriers.

## Benefits of data lakes

### Save on ingestion costs

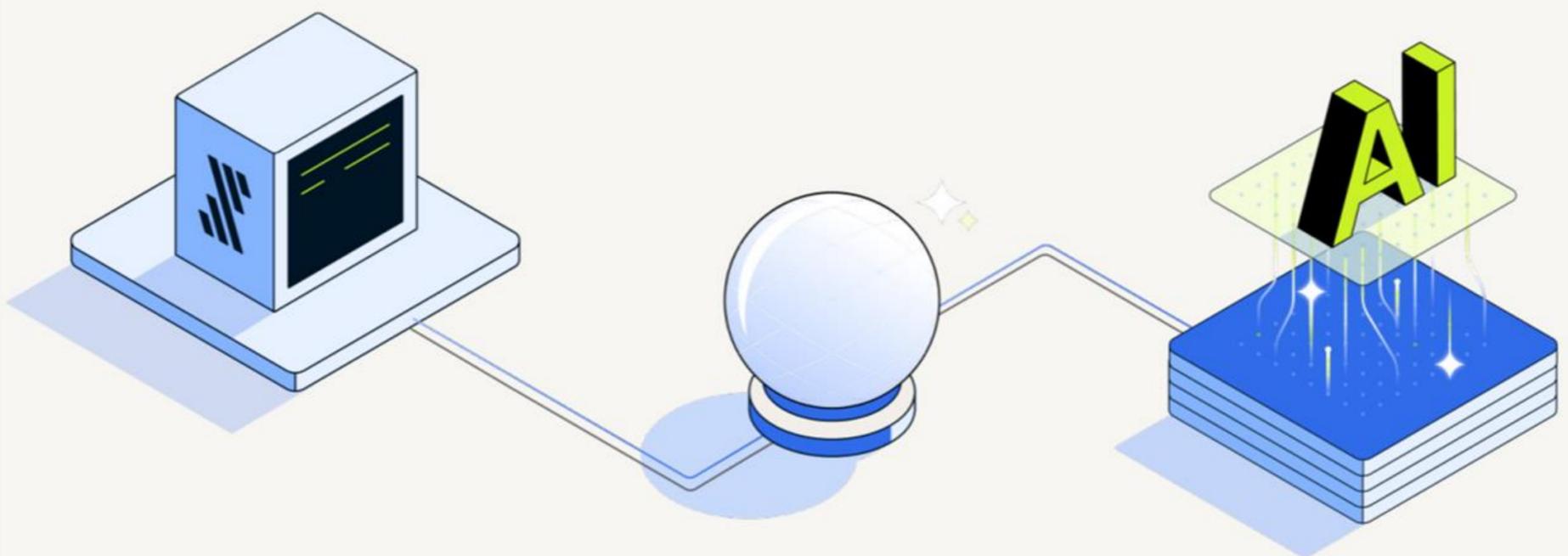
Minimize or remove the need to duplicate data across multiple analytic compute engines

### Future proof your data

Open table formats and API-based catalogs allow you to load your data once to the lake and query as needed, making data more accessible

# Impact of data lakes on AI innovation

Data lakes are increasingly central to AI innovation due to their ability to store and process vast volumes of structured, semi-structured, and unstructured data. This flexibility is critical for generative AI, which thrives on diverse and context-rich datasets.



## GenAI

Generative AI models, particularly large language models (LLMs), rely on centralized, fresh, and governed data to produce accurate and relevant outputs.

## Universal storage

Data lakes support this by serving as a universal storage layer that feeds AI/ML workflows, allowing for exploratory analytics and predictive modeling without duplicating datasets.

## RAG applications

The ability to integrate proprietary data into foundation models using retrieval-augmented generation (RAG) architectures enhances the specificity and utility of AI applications across industries.

# Common challenges for data lake adoption

Adopting a modern data lake can unlock powerful opportunities for AI and analytics, but organizations often encounter three critical hurdles along the way. Addressing these three challenges is essential to accelerate time to market and avoid common pitfalls in data lake adoption.

01

Overcoming legacy data infrastructure limitations

02

Ensuring data quality & pipeline reliability to address fragmented systems

03

Simplifying purchase and implementation processes

Companies like Tinititi show how adopting scalable data lakes makes a difference. Using Fivetran Managed Data Lake Service, onboarding gets faster, operational overhead drops, and teams get real-time insights for AI-powered marketing and analytics.

# Fivetran helps Tinuiti build a scalable data lake for AI-driven marketing

Tinuiti is the largest independent full-funnel performance marketing agency in the U.S., managing nearly \$4 billion in media spend across major digital channels. The company is known for its proprietary measurement product, Bliss Point by Tinuiti, which uses advanced data science—including causal inference and machine learning—to deliver real-time marketing insight.

## Challenge

Tinuiti's data warehouse limited control and scalability, causing slow onboarding and high operational overhead. Fragmented platforms and inconsistent schemas led to unreliable data ingestion and manual pipeline maintenance. Engineers spent excessive time maintaining connectors, delaying customer onboarding and business growth.

## Solution

With the help of Fivetran, Tinuiti transitioned to an Amazon S3 data lake with full control, scalability, and flexibility.

Fivetran automated ingestion from 30+ sources and standardized data into Delta Lake and Iceberg open table formats for interoperability.

Data was cataloged in AWS Glue, improving governance, security, and query performance.

## Outcome

Tinuiti now serves all customers from a single, centralized platform.

- Client onboarding accelerated 120x by reducing data ingestion pipeline setup time from 2-4 weeks to under an hour.
- Manual data lake and pipeline maintenance were eliminated, saving engineers significant time.
- Standardized data accelerated time-to-value and enabled real-time, trustworthy insights for global campaigns.



Fivetran's Managed Data Lake Service gives us the flexibility to direct data anywhere we want and consume it any way we want. We can choose what to bring into the warehouse or send elsewhere, which has reduced our costs, increased security, and improved our data governance."

— Lakshmi Ramesh, VP of Data Services at Tinuiti

[Read the full case study here](#)

Let's dive deeper into how Fivetran Managed Data Service can help you address your own similar challenges.



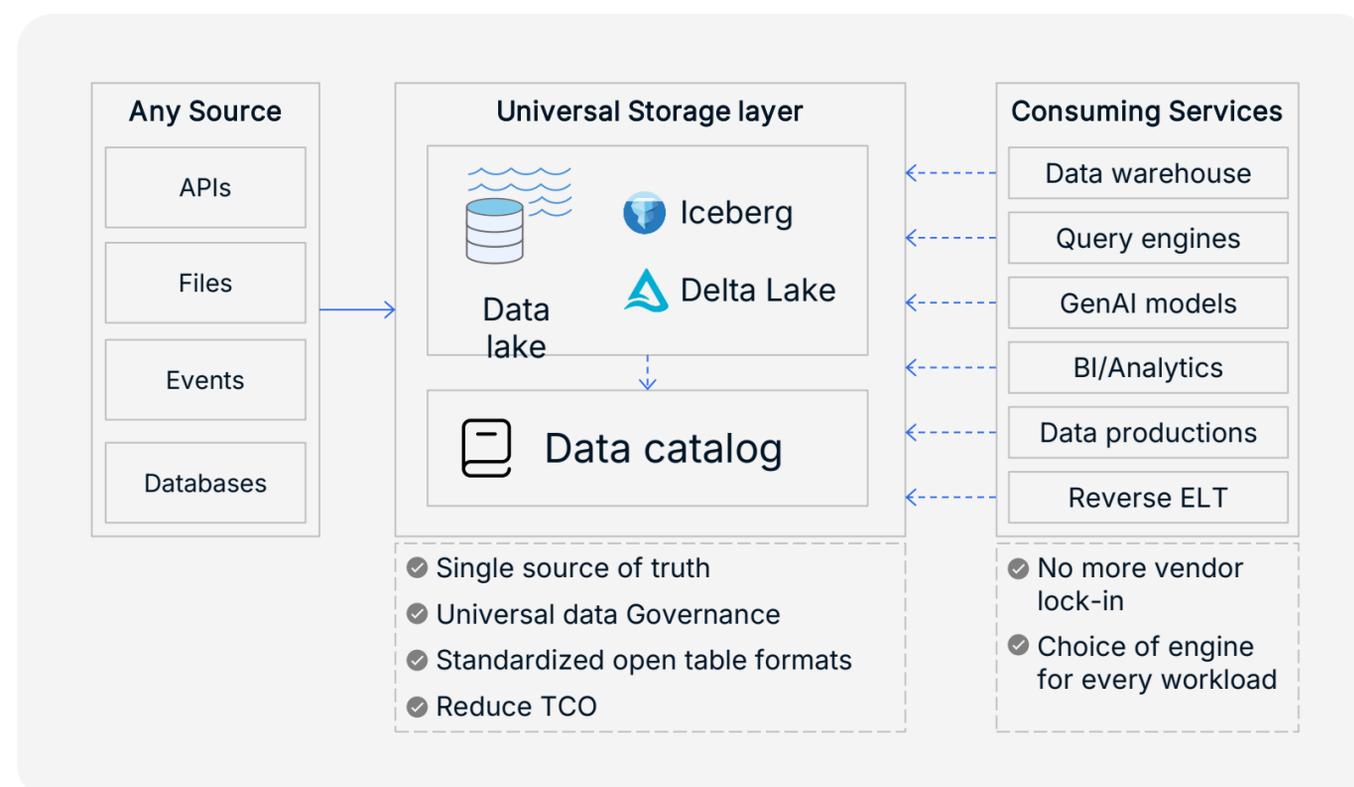
# Overcoming legacy data infrastructure limitations

Legacy data infrastructure causes modernization and scalability issues for existing operations and can create uncertainty when experimenting with AI. Without data integration you can trust, how do you know your data is compliant, accurate, and actionable?

Simplifying data architecture is made easy with Fivetran Managed Data Lake Service. This service consolidates data from various sources into a single, structured foundation by enabling a universal data storage layer. This centralization allows enterprises to write data once and read it many times—helping ensure compliance, stability, and scalability. It also allows organizations to locate and delete specific records to adhere to regulations like GDPR.

Data consolidation into a data lake allows organizations to scale their data infrastructure without the need to repeatedly stand-up new databases or warehouses. This flexibility is particularly valuable during periods of growth or acquisitions, as new lines of business can be quickly onboarded into the interoperable data environment.

Fivetran's platform integration with open table formats like Delta Lake and Apache Iceberg further enhances this architecture by applying structure to previously unstructured data lakes. This results in transactionally correct, query-ready data that supports both operational analytics and advanced AI/ML use cases.



## Security is top of mind

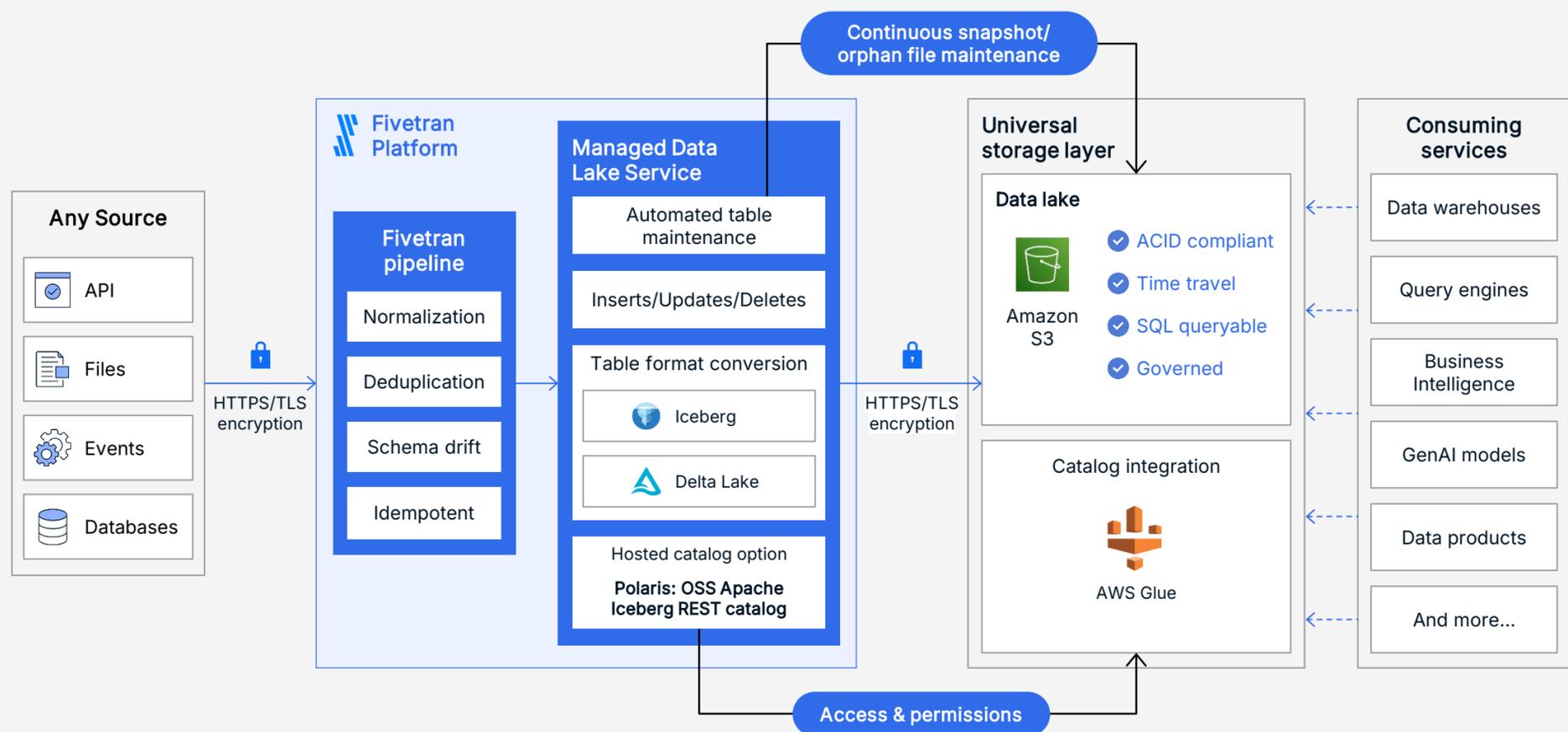
Fivetran Managed Data Lake Service encrypts data in transit and at rest—blocking or hashing sensitive data before it enters a destination. When hashing data, Fivetran returns an anonymized value, so the data remains queryable.

# Ensuring data quality and pipeline reliability

Data quality and pipeline reliability are often results of a fragmented data environment. Not only does Fivetran Managed Data Lake Service consolidate and structure data, it also eliminates the burden of manual schema evolution and pipeline maintenance.

To ensure data quality, Fivetran Managed Data Lake ingestion process includes normalization, standardization, deduplication, and automatic schema migration. Through regularly performed maintenance operations, Fivetran maintains an efficient data storage environment, ensuring data in the lake is accurate even as it changes upstream.

The focus on connector quality and high-performance pipelines allows Fivetran to ensure organizations have access to their business-critical data. Through continuous monitoring and automatic handling of merges, deletes and schema changes, Fivetran enables reliable, hands-off data integration that maintains quality without manual oversight. This helps ensure uninterrupted data flow and consistent performance, even across complex environments. With real-time access and automated governance, Fivetran empowers organizations to scale AI initiatives without compromising data integrity or operational efficiency.



# Simplifying software purchasing and implementation

Many organizations are hindered by lengthy procurement cycles. Additionally, complicated implementation processes can delay AI projects and increase costs, making it difficult for organizations to leverage AI effectively.

AWS Marketplace streamlines complex buying processes with predefined contracts—eliminating the need for lengthy negotiations and time-consuming procurement processes.

Fivetran also offers a 14-day free trial, allowing customers to experiment with the platform quickly and cost-effectively.

This streamlined approach is particularly valuable for data teams with limited budgets or tight timelines, enabling them to move quickly and efficiently. As a trusted entry point, AWS Marketplace provides technical decision makers access to Fivetran, reducing friction and lowering barriers to adoption.

Fivetran further reduces implementation complexity of data pipeline setup by automating ingestion from 700+ sources. In addition to pre-built connections, Fivetran's Connector SDK offers automated ingestion from custom-built sources, reducing the time and resources needed to operationalize data infrastructure.

Together, Fivetran and AWS Marketplace free data teams to focus on strategic initiatives, accelerating time-to-value and improving ROI.



Two things are great about Fivetran's flexibility and cost savings. The initial load is free and every connector is free to use for 14 days. We can experiment with data, make business decisions and empirically forecast costs, allowing us to focus on high-value analysis tasks at very low risk."

— Sandro Frattura, Analytics Engineering Manager, HubSpot People Operations

## Better together with Fivetran + AWS

- Enable seamless compatibility with existing cloud infrastructures by integrating with AWS services like Amazon S3 and AWS Glue
- Make data query-ready and compliant with standards like ACID and GDPR by automating ingestion into open table formats like Apache Iceberg and Delta Lake
- Scale analytic and AI initiatives with tools like Amazon SageMaker—using data processed and maintained by Fivetran
- Enhance discoverability and governance through integrations with catalogs such as AWS Glue and Databricks Unity Catalog



• Amazon Redshift Ready  
• AWS Marketplace Seller  
• Data & Analytics Software Competency

# Conclusion

The evolution of data lakes has revolutionized the way organizations manage and utilize their data, particularly in the realm of AI and machine learning. Fivetran Managed Data Lake Service stands out as a pivotal solution, addressing the critical challenges of legacy data infrastructure, data quality, and complex procurement and implementation processes. By automating ingestion and transformation, ensuring consistency, and simplifying procurement through AWS Marketplace, Fivetran empowers organizations to build scalable, reliable, and governed data lakes.

These advancements enable organizations to leverage their data for advanced analytics and AI-driven decision making, driving innovation and competitive advantage. Tinuiti's success highlights how Fivetran provides tangible benefits that enhance operational efficiency and accelerate AI adoption.

As you explore the potential of data lakes and AI, consider how Fivetran Managed Data Lake Service can transform your data infrastructure, streamline your processes, and unlock new opportunities for growth and innovation.

With Fivetran, you can confidently navigate the complexities of modern data lake management and harness the full power of your data to drive meaningful business outcomes.

[Check out Fivetran in AWS Marketplace for a demo and trial](#)

## About Fivetran

Fivetran, the global leader in data movement, helps customers use their data to power everything from AI workloads and ML models to predictive analytics and operational outputs. The Fivetran platform reliably and securely centralizes data from hundreds of databases and SaaS applications into any destination — whether deployed on-premises, in the cloud, or in a hybrid environment. Thousands of global brands, including Autodesk, Condé Nast, JetBlue, and Morgan Stanley, trust Fivetran to move their most valuable data assets to fuel analytics, drive operational efficiencies, and power innovation. For more info, visit [fivetran.com](https://fivetran.com).

