Enterprise Strategy Group™ by TechTarget

WHITE PAPER

Automated Data Movement for Modern Organizations

Fivetran for AWS Marketplace

By Justin Boyer, IT Validation Analyst; and Tony Palmer, Principal Analyst
Enterprise Strategy Group

November 2023

Contents

Abstract

This White Paper from TechTarget's Enterprise Strategy Group documents our detailed evaluation of Fivetran for Amazon Web Services (AWS). We evaluated how Fivetran provides automated data movement across cloud data platforms.

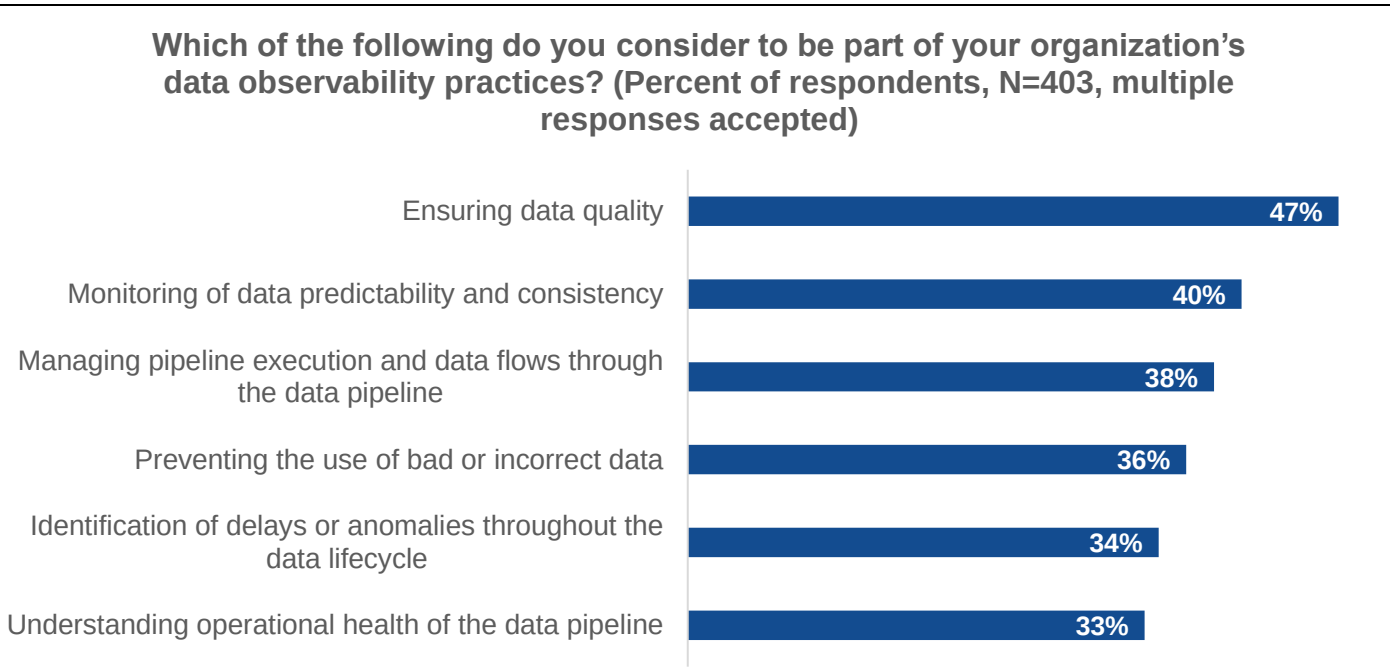
Challenges

As data continues to grow in organizations, there is an increasing need to move the data from the place of origin to places where analysis can be made. The variety of data and sources makes managing data movement challenging as organizations grow. Traditional extract, transform, and load methods require weeks or months to complete.

Moving large amounts of data is made more challenging due to data quality issues, data silos, and the volume of data. According to Enterprise Strategy Group research, data quality is the most commonly cited investment priority over the next 12–18 months for survey respondents.[1]

In addition, Enterprise Strategy Group asked respondents in a separate research survey about their organization's data observability practices. As shown in Figure 1, ensuring data quality was the most commonly cited data observability practice, followed by monitoring data predictability and managing the data pipeline.[2]

Figure 1. Top Six Tasks Considered Part of Data Observability Practices



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

Organizations need a data movement solution that reduces complexity for data engineers, connects to the many data sources used by modern companies, and maintains data reliability and security.

[1] Source: Enterprise Strategy Group Complete Survey Results, *2023 Technology Spending Intentions Survey*, November 2022.
[2] Source: Enterprise Strategy Group Complete Survey Results, *The State of DataOps*, August 2022.

Automated Data Movement with Fivetran for AWS Marketplace

Fivetran is an automated data movement platform that reduces complexity for data engineers. It's a no-code, self-service platform with over 400 connectors to multiple categories of data sources. These allow data engineers to build pipelines in minutes.

Fivetran also has Lite connectors available through a By Request program. With this option, customers can request a Lite connector for any SaaS application and Fivetran will develop them in a fraction of the time of standard connectors—while still backed by the same high-quality data integrity and security. When Fivetran doesn't have a pre-built connector for a source or isn't building a Lite connector, their Custom Connectors can automate data movement as well. Customers write a cloud function to extract the data from their source, and Fivetran handles the loading and transforming into the destination.

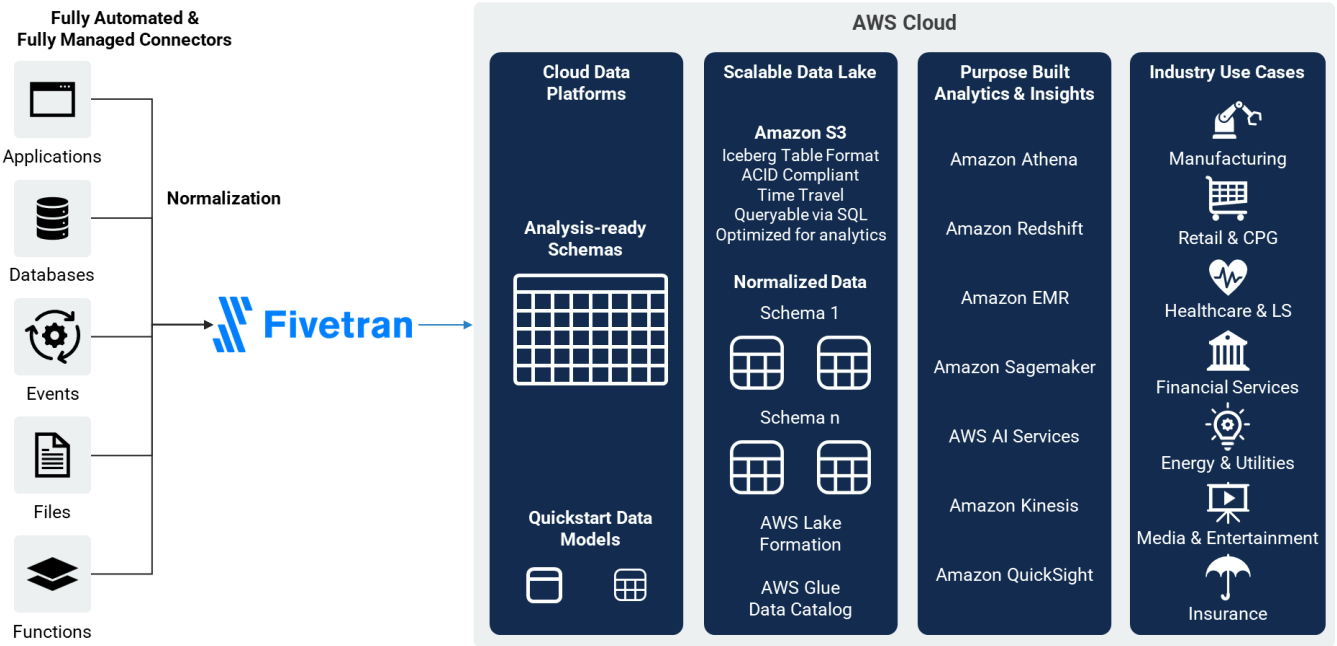
Fivetran provides a SaaS platform, fully managed by over 300 engineers with 99.9% uptime. Creating data pipelines is declarative, and once data engineers design the pipeline, Fivetran handles the maintenance. There is zero effort required by the customer for ongoing ops.

In addition to the SaaS offering, Fivetran offers a Local Data Processing deployment option for relational database sources and SAP that can be hosted within a customer's infrastructure (on-premises, in a private cloud, or in a public cloud) providing a range of deployment options for critical data sources.

Data security and compliance are serious concerns for modern organizations when moving data. With Fivetran, security and compliance are built in. Encrypted connections protect data in transit. Fivetran blocks or hashes sensitive data before it reaches the destination. Full logging and role-based access control (RBAC) are included to provide complete control over how data is handled and by whom.

Fivetran provides the tools to centralize and democratize data, modernize infrastructure, and build high-value data solutions.

Figure 2. Fivetran Automated Data Movement



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

Enterprise Strategy Group Validated

Enterprise Strategy Group examined how Fivetran Data Pipelines in AWS Marketplace enables organizations to focus on analytics rather than data engineering by automatically ingesting and centralizing data from hundreds of data sources into a schema that is ready for analysis.

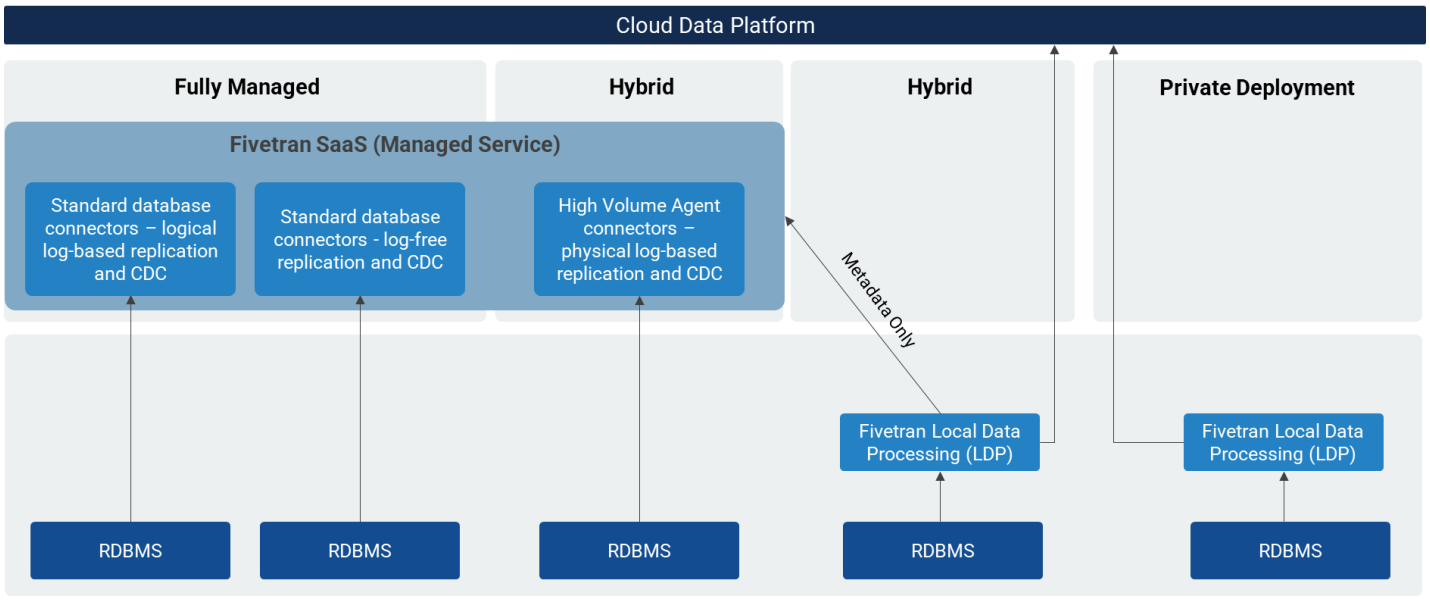
Database Connectors

First, we looked at Fivetran's capabilities around database connectors. As organizations grow, they'll inevitably start building up a large amount of data in several different databases and applications. Fivetran offers more than 400 distinct data sources, including databases, applications, object stores, and more.

Figure 3 outlines Fivetran's data replication options available to customers. Fivetran offers a fully managed SaaS service as well as hybrid and self-hosted options. There are multiple replication options. Log-based replication—which can be logical or physical—uses transaction log files as the source for changes, while log-free replication directly queries the source database tables for changes. There are two hybrid models" Fivetran SaaS High Volume Agent (HVA) hybrid where the agent is on the source system, and Fivetran local data processing (LDP) hybrid where the agent is at or near the source and the LDP infrastructure is customer-managed behind their firewall on-premises, in a private cloud, or in a public cloud. Fivetran SaaS HVA and Fivetran LDP provide the ability to do high-volume replication using filtering, compression, and encryption.

Some organizations may require an air-gapped solution due to the sensitive nature of the data they process, such as financial services or healthcare companies. Fivetran's LDP self-hosted private deployment option provides fast speed using the agent. This option doesn't send any data to the Fivetran SaaS platform but communicates only with the destination platform. The same filtering, compression, and encryption apply as with the Fivetran LDP hybrid option, but with an air-gapped environment controlled by the customer.

Figure 3. Database Connectors



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

Fivetran also features many built-in data security, governance, and compliance measures. Organizations can choose among several secure connection options, such as SSH, VPN, or AWS PrivateLink. Fivetran allows organizations to block any source data that shouldn't move to the cloud data platform and anonymize PII

(personally identifiable information) data via hashing. Data that is processed in either the Fivetran SaaS or LDP service is purged once it's successfully written to the destination.

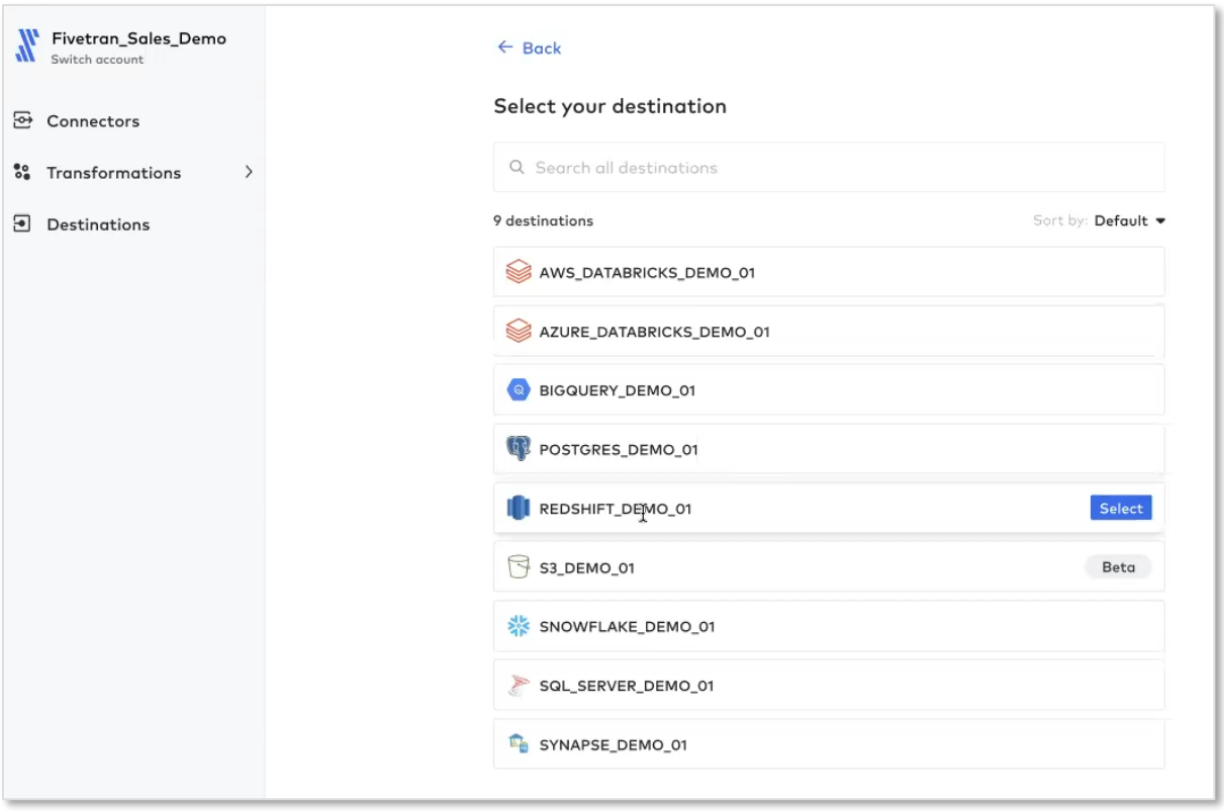
Organizations can control how data is handled, and by whom, using data access controls and logging for visibility and auditing, along with RBAC and programmatic automation of user, role, and management policies via REST API allow organizations to control how data is handled and by whom.

Fivetran's service encrypts all data in flight (using Transport Layer Security) and at rest (using 256-bit Advanced Encryption Standard), which includes any data that passes ephemerally though the Fivetran service.

Fivetran Deployment

Enterprise Strategy Group walked through a deployment of the Fivetran SaaS service. The first step was to select a destination. The test environment had nine destinations configured, but there is no predefined limit. We chose Amazon Redshift.

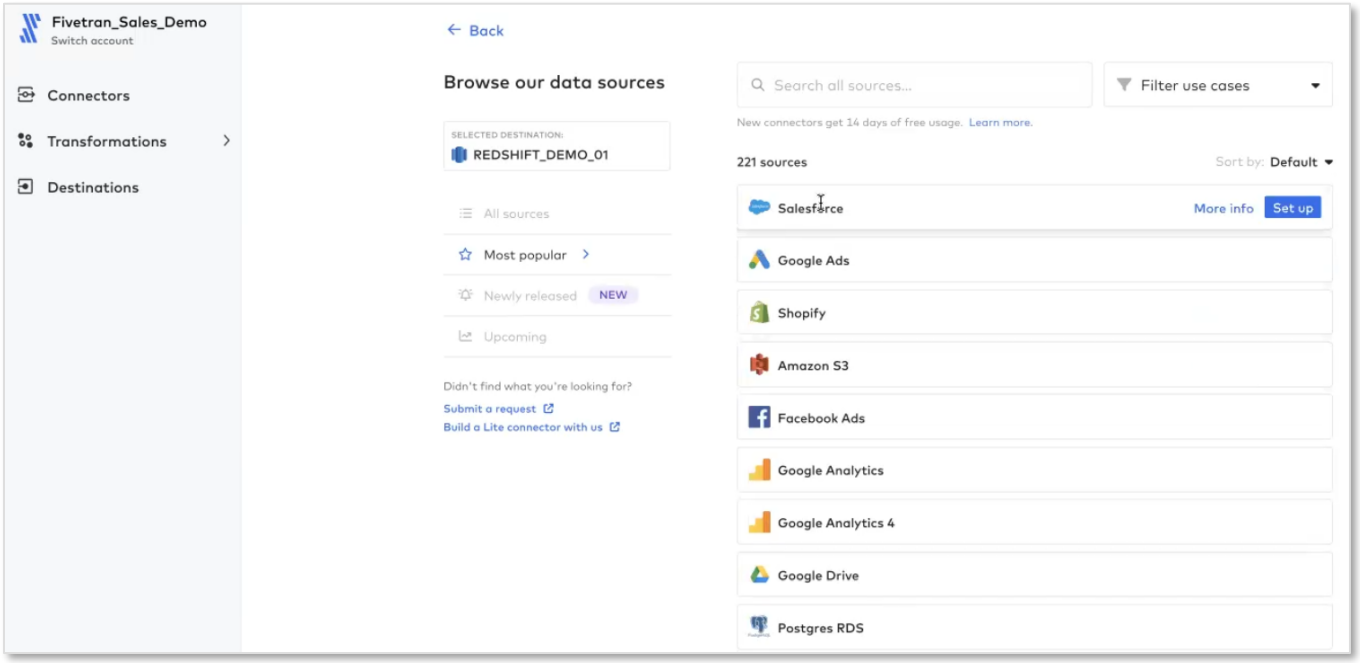
Figure 4. The Fivetran User Interface—Selecting a Destination



Source: Fivetran

Next, we selected a data source. As of this writing, Fivetran supports more than 400 distinct data sources, including databases, applications, object stores, and more. For this test, we selected Salesforce.

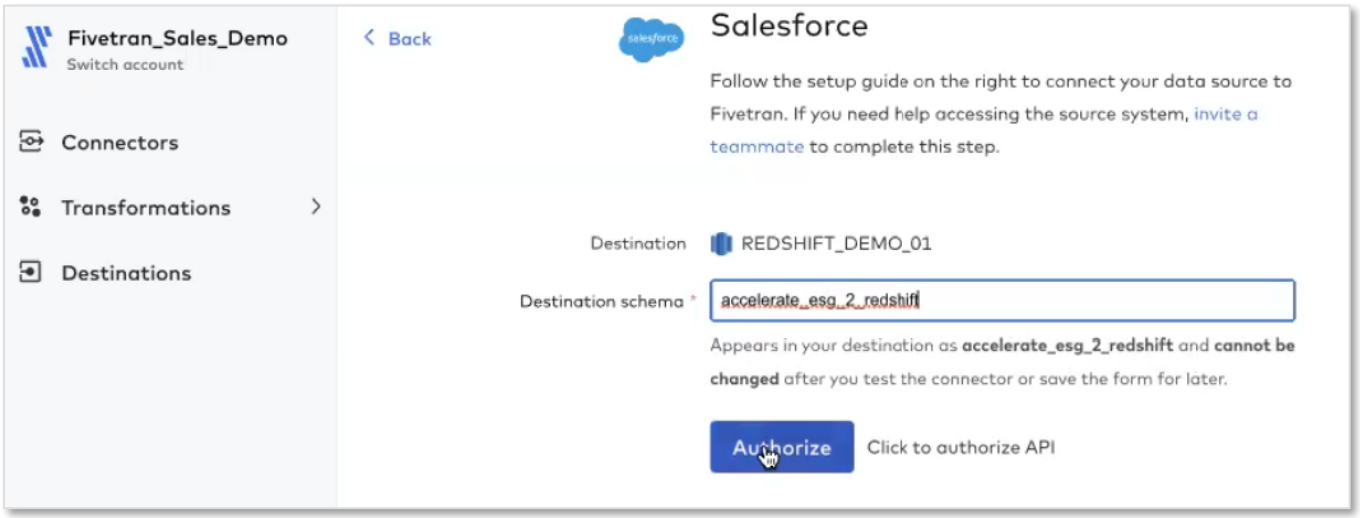
Figure 5. Selecting a Data Source



Source: Fivetran

Next, we named the destination schema. It's important to note that users don't have to create a schema ahead of time in Fivetran SaaS. Fivetran manages the schema creation in the destination and also handles any schema drift over time. When we clicked *Authorize*, and Fivetran authenticated us with the Salesforce source instantly. We clicked *Save and Test*, and Fivetran executed multiple connectivity tests, which were completed in about 10 seconds. Following this, Fivetran was able to access the Salesforce schema and allowed us to select the schemas, tables, and columns from the Salesforce source.

Figure 6. Authenticating with the Data Source



Source: Fivetran

Next, we selected the data to import. Users can select every table or as few as are needed. It's important to note that Fivetran gives the option to manually identify and anonymize sensitive data in any column with a click. Fivetran is planning to release an auto-PII detection feature soon for even more proactive data privacy and anonymization.

Figure 7. Selecting Data

Source: Fivetran

The last step before starting synchronization was to specify how Fivetran should handle changes to the schema. Organizations can allow all new schemas, tables, and columns; allow new columns only; or block all changes.

The entire process, from creating the connector to starting synchronization, took less than five minutes. Traditionally, a team would need to run through the design, build, document, test, create, monitor, and manage phases that often require multiple weeks to complete.

Following the historical sync, orchestration of change data capture (CDC) and scheduling is built in to the platform. Synchronization is customizable. Fivetran SaaS standard connectors support five-minute sync, and Fivetran SaaS HVA connectors support one-minute sync. The Fivetran LDP platform is a continuous sync process and doesn't require setting a sync time.

Fivetran SaaS has more than 100 Quickstart transformations that enable fully automated data aggregations and transformations though pre-built models that can get an organization started rapidly on its desired data outcomes. After establishing the foundation, organizations can then build and customize transformations and enrich data sets as needed, without having to start from scratch.

Conclusion

Organizations are increasingly distributed, and they need to combine data from dissimilar sources and formats across multiple locations, adding complexity and cost to the access, management, and securing of data. This creates multiple challenges. Organizations' limited data visibility across hybrid locations makes managing, accessing, and extracting value from their data more difficult and time-consuming, resulting in negative impacts on both time to production and time to innovation. Organizations want to use their data in multiple ways for numerous purposes, including leveraging AI/ML for critical decision-making, optimizing operations, developing new revenue streams, and enhancing the customer experience. Enterprise Strategy Group research supports these ideas, as the top three investment areas for survey respondents for the next 12-18 months are data quality, business intelligence, and data integration.[3]

Fivetran attempts to address these challenges by providing a platform to extract, load, and transform data from various sources into organizations' data warehouses with speed and security. It seeks to help organizations quickly ramp up their data transformation initiatives by eliminating the long development cycles often associated with data transformation projects. Fivetran's no-code, self-serve model, wealth of connectors and data sources, and flexible deployment options work together to reduce complexity, duplication of effort, errors, and costs.

Enterprise Strategy Group validated Fivetran's ability to enable fast and reliable data consolidation and centralization of data to provide true self-service analytics. With Fivetran, organizations can extract insights across a consolidated dataset, eliminating individual silos of data. They can save time using the more than 100 Quickstart transformations featuring pre-built data models for various sources. During our testing, we deployed Fivetran to connect Salesforce to Amazon Redshift in less than five minutes. The data is secured via encryption in transit and at rest, throughout the entire data transfer process, ensuring security and compliance with industry regulations. Once an organization creates a data pipeline, Fivetran takes over ongoing CDC from the source, ongoing operations, and maintenance.

Accurate data is more valuable to businesses than ever, making mistakes more costly as well. If your organization needs to consolidate, transform, and move data without long development cycles, expensive maintenance, and costly errors, Enterprise Strategy Group suggests you take a serious look at the Fivetran Data Pipelines solution in AWS Marketplace.

AWS Marketplace

AWS Marketplace is a curated digital catalog that organizations can use to find, buy, deploy, and manage third-party software, data, and services that they need to build solutions and run their businesses. AWS Marketplace includes thousands of software listings from popular categories such as security, networking, storage, machine learning, IoT, business intelligence, database, and DevOps. Additionally, organizations can browse and subscribe to data products. AWS Marketplace makes it easy to launch pre-configured software with just a few clicks and choose software solutions in Amazon Machine Images (AMIs), software as a service (SaaS), as well as other formats. AWS Marketplace also simplifies software licensing and procurement with flexible pricing options which include free trial, hourly, monthly, annual, multi-year, and a Bring Your Own License (BYOL) model.

[3] Source: Enterprise Strategy Group Complete Survey Results, *2023 Technology Spending Intentions Survey*, November 2022.

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