

Best practice guide

Moving data and powering innovation with Fivetran and Databricks

Accelerate data movement into the Databricks Data Intelligence Platform to execute analytic workloads, AI and GenAI



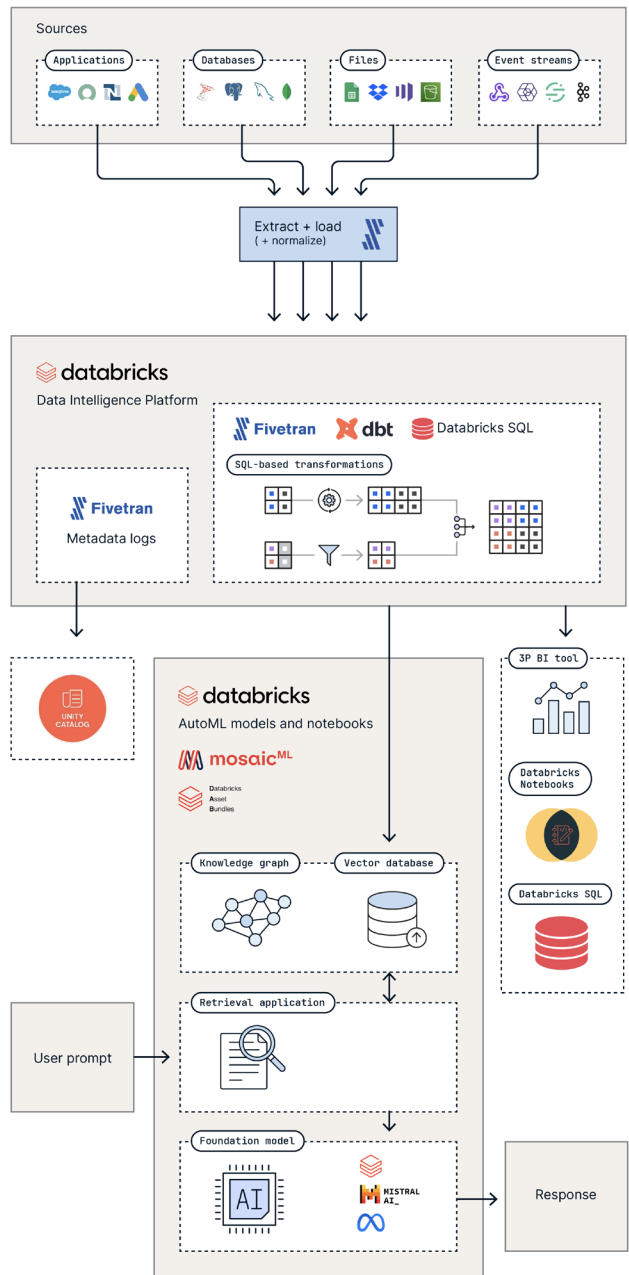
Data integration is deceptively complex. Considerations include scoping and scaling infrastructure, ensuring availability, recovering from failures and maintaining the system in response to changing data sources and business needs. Many common data integration tools offer frameworks for solving these tasks but still demand a considerable degree of engineering work from end users.

The Fivetran automated data movement platform automatically, reliably and securely centralizes data from hundreds of SaaS applications, files, events and databases into the Databricks Data Intelligence Platform. By combining Fivetran and Databricks, you gain the following capabilities:

Automated, zero-maintenance data pipelines	Secure data replication for the Databricks Data Intelligence Platform	Simplified data management for all analytics and AI use cases
◆ Modern ELT approach: Fivetran automatically loads normalized data into Databricks ◆ Continuous, fast data delivery with minimal configuration ◆ 500+ natively-supported connectors for industry standard databases, SaaS apps, event streams and files	◆ Log-based CDC replication delivers high data volumes in real time to Databricks without impacting source systems ◆ Teams can extract and combine business-critical data for a 360° perspective ◆ Rigorously tested for compliance with common regulatory and industry standards	◆ Replace data silos and unify data teams on a single platform with a modern data stack built on Fivetran and Databricks ◆ Use the Databricks platform to store structured, semi-structured and unstructured data

How Fivetran and Databricks work together

Fivetran seamlessly integrates with the Databricks ecosystem, performing a critical role in moving data into a data lake or warehouse. Databricks provides a destination for the data and tools to assist with governance (Unity Catalog), development (Asset Bundles) and AI capabilities (AutoML and Mosaic AI).



Using Fivetran and several tools in the Databricks ecosystem, you can follow a straightforward workflow for data applications of all kinds.

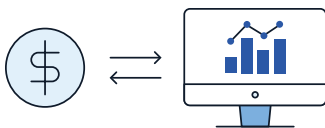
For conventional analytics and business intelligence, you can:

- 1 Use **Fivetran** to integrate data into the **Databricks** Data Intelligence Platform
- 2 Use **Fivetran**, **dbt** and/or **DatabricksSQL** to transform data into analytics-ready data models.
- 3 Use **Databricks SQL** or **Databricks Notebooks** to produce reports and dashboards.

For AI/ML applications, you can:

- 1 Use **Fivetran** to integrate data into the **Databricks** Data Intelligence Platform
- 2 Use **Fivetran**, **dbt** and/or **Databricks SQL** to transform data into analytics-ready data models.
- 3 Use **AutoML** or **MosaicML** to set up ML or RAG (retrieval-augmented generation) models
 - ◆ **AutoML** natively supports off-the-shelf machine learning models for classification, regression and forecasting, and automates tuning and testing.
 - ◆ You can also embed data as vectors into the vector database of your choice with **MosaicML**.
- 4 Use **Databricks Asset Bundles** to organize the projects for your application using the languages and libraries of your choice, reducing development overhead.

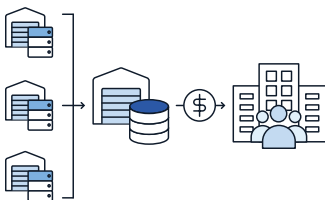
This combination of platforms and tools enables you to build the full range of analytics and AI/ML applications:



Curated, white-glove, real-time analytics

Data science is a difficult profession and talent is scarce. You can provide "insights as a service" for customers who choose to forgo building analytics capabilities in-house.

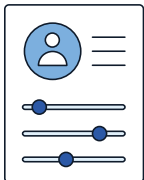
- ◆ **Custom reporting and dashboards:** Create tailored reports and interactive dashboards for clients, allowing them to visualize their key metrics and make data-driven decisions.
- ◆ **Predictive maintenance:** Offer real-time monitoring and predictive analytics for industries like manufacturing and energy, helping them reduce downtime and optimize maintenance schedules without the need for a dedicated analytics team.



Data aggregation and sharing services

You will gain valuable insights by collecting data from across a market, industry, supply chain, etc.

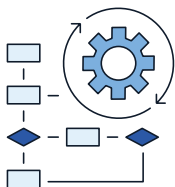
- ◆ **Market intelligence platforms:** Aggregate data from multiple sources to provide comprehensive market insights, helping businesses stay ahead of trends and make informed strategic decisions.
- ◆ **Collaborative research databases:** Enable academic and corporate researchers to share and access vast datasets, fostering innovation and data discovery.



Personalization and recommendation

You can use behavioral, psychographic and other data for decision support. This approach can enhance customer experiences through tailored marketing strategies, personalized content delivery and product recommendations based on individual user preferences and past interactions. Some examples include:

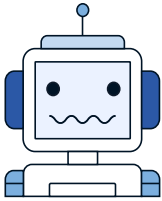
- ◆ **E-commerce recommendations:** Using purchase history, browsing behavior and customer preferences to suggest products that are more likely to interest individual users, increasing sales and customer satisfaction.
- ◆ **Content personalization:** Tailoring news, articles and multimedia content to individual user preferences, increasing engagement and retention on digital platforms.



Automation

You can use data to trigger all kinds of automatic behavior within your organization's operations including automating routine tasks to improve efficiency, using predictive maintenance to reduce downtime, and deploying intelligent chatbots for customer service to handle inquiries swiftly and effectively, as well as:

- ◆ **Customer service automation:** You can use chatbots to handle routine customer inquiries, improving response times and freeing up human agents for more complex issues.
- ◆ **Supply chain optimization:** Inventory management automation, order processing and logistics planning using real-time data ensures can safeguard deliveries and reduce operational costs.



Generative AI of all kinds

Generative AI has the potential to become a phenomenally powerful productivity aide as well as the backbone for a wide range of innovative products, including:

- ◆ **Content creation:** Automate the generation of written content, graphics, videos and music, helping creative professionals enhance their productivity and experiment with new forms of artistic expression.
- ◆ **AI-driven design:** Use AI to help engineers and designers create complex designs and prototypes more efficiently.
- ◆ **Personalized learning and education tools:** Develop educational tools that adapt to the learning pace and style of each student, offering personalized tutorials, exercises and feedback to maximize learning efficiency.
- ◆ **Interactive entertainment:** AI will almost undoubtedly power more and more next-generation video games, virtual reality and augmented reality experiences. AI can dynamically generate content, story arcs and responsive, immersive environments.

In short, with Fivetran and Databricks, the world of data is your oyster. The types of data applications you can build are limited only by your imagination.

For functionality not directly supported by the Databricks ecosystem, many third-party providers partner with and seamlessly integrate with the Databricks Data Intelligence Platform.

Getting started with Fivetran and Databricks

There are two ways to get started with Fivetran:

- 1 Through **Databricks Partner Connect** via your Databricks dashboard.
Recommended!
- 2 Through fivetran.com.

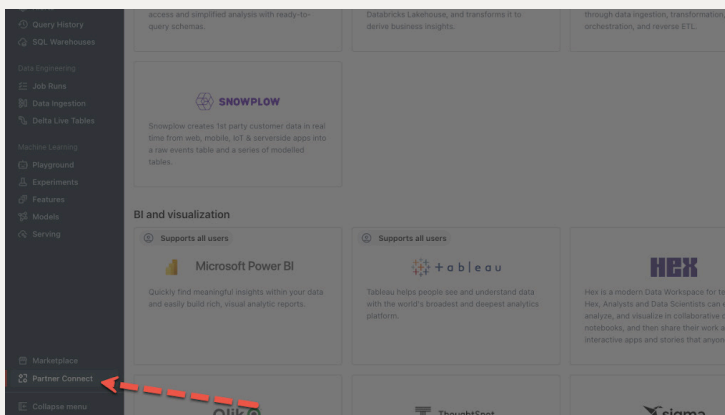
Here are step-by-step instructions for getting started with Fivetran through Databricks Partner Connect:

STEP 1

Log in to your Databricks account.

STEP 2

From the **Dashboard**, select **Partner Connect** and search for **Fivetran**.



STEP 3

Select **Fivetran** and connect. You can choose the specifications for your Databricks destination. Click **Next**.

STEP 4

For **Email**, enter the email address that you want Fivetran to use to create a 14-day trial Fivetran account for you, or enter the email address for your existing Fivetran account. Click the button with the label **Connect to Fivetran** or **Sign in**. If needed, you can also review some connection details. Click **Connect to Fivetran**.

The screenshot shows the Fivetran connection configuration window. The 'Email' field is at the top. Below it, the 'Connection details' section is expanded, showing fields for 'User' (FIVETRAN_USER), 'Personal access token' (masked with asterisks), 'Server Hostname' (dbc-d1905111-ae7a.cloud.databricks.com), 'Port' (443), and 'HTTP path' (/sql/1.0/warehouses/09025939304721). At the bottom, there is a disclaimer about data processing and a 'Connect to Fivetran' button.

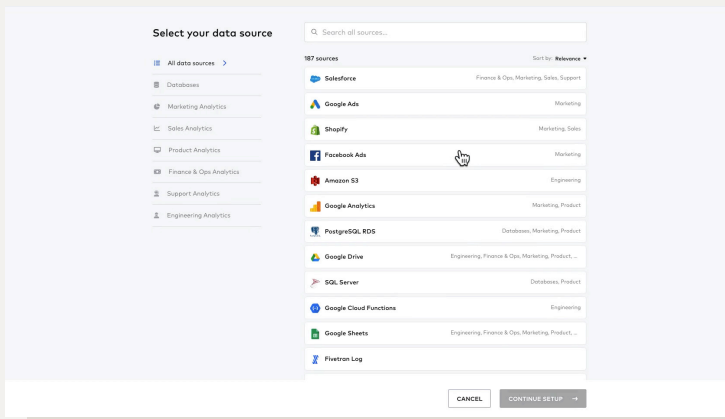
STEP 5

After activation, you will be prompted to create a password before selecting **Start Free Trial**. No credit card is required to sign up for the free trial and use Fivetran for 14 days, and the free trial only begins once you have successfully completed an initial sync.

The screenshot shows the 'Create your password to start your free 14-day trial with Fivetran' screen. It includes a blue sidebar with the text 'Start replicating your data to your Databricks account now - for free.' and a main white area with a password creation form. The form has two password input fields (the second with a strength indicator), an optional phone number field, and a 'Start Free Trial' button. A disclaimer at the bottom states: 'Creating a password signifies that you have read and agreed to our Terms of Service and Privacy Policy.'

STEP 6

After creating an account, Partner Connect brings you to the Fivetran UI, where you can add connectors and have all of your data at your fingertips. After choosing a connector, follow the on-screen instructions to set it up.



Once you have created and synced your first connector, your free 14-day trial starts. Fivetran will deliver the data you need in a timely manner and in an easy-to-query format.

To see how you can further combine Fivetran with Databricks, consider attending or watch a recording of a hands-on lab. It can take as little as a few lines of SQL and Python to start spinning up your organization's next breakthrough.



Fivetran, the global leader in data movement, helps customers use their data to power everything from AI applications and ML models, to predictive analytics and operational workloads. The Fivetran platform reliably and securely centralizes data from hundreds of SaaS applications and databases into any cloud 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.

©2024 Fivetran Inc.