

CONTENIDO

1. INTRODUCCIÓN.....	3
REFERENCIAS	4
ACCESO A LA HERRAMIENTA.....	4
2. OVERVIEW DE LA HERRAMIENTA	5
3. CASO PRÁCTICO.....	6
4. VENTAJAS/DESVENTAJAS	19
5. CONCLUSIÓN	20

1. INTRODUCCIÓN

Fivetran es una potente herramienta de **orquestación de datos** diseñada para simplificar y acelerar el proceso de consolidación de información proveniente de diversas fuentes en una única ubicación centralizada. Fundada en 2012, Fivetran se ha convertido en una solución líder en el espacio de la integración de datos, ofreciendo una plataforma robusta y fácil de usar para empresas de todos los tamaños.

Esta especializada en la replicación e integración de datos que facilita la conexión entre fuentes de datos de diversos tipos y formatos con destinos de cualquier tipo incluyendo Bases de Datos, SAP, Data Warehouse y Data Lakes.

Tiene varios conectores predeterminados para las transformaciones, pero su principal característica reside en la fácil unión con [DBT](#), ya que los procesos de ETL se crean mediante DBT y para la ejecución de ellos se utiliza Fivetran con una interfaz más intuitiva y fácil.

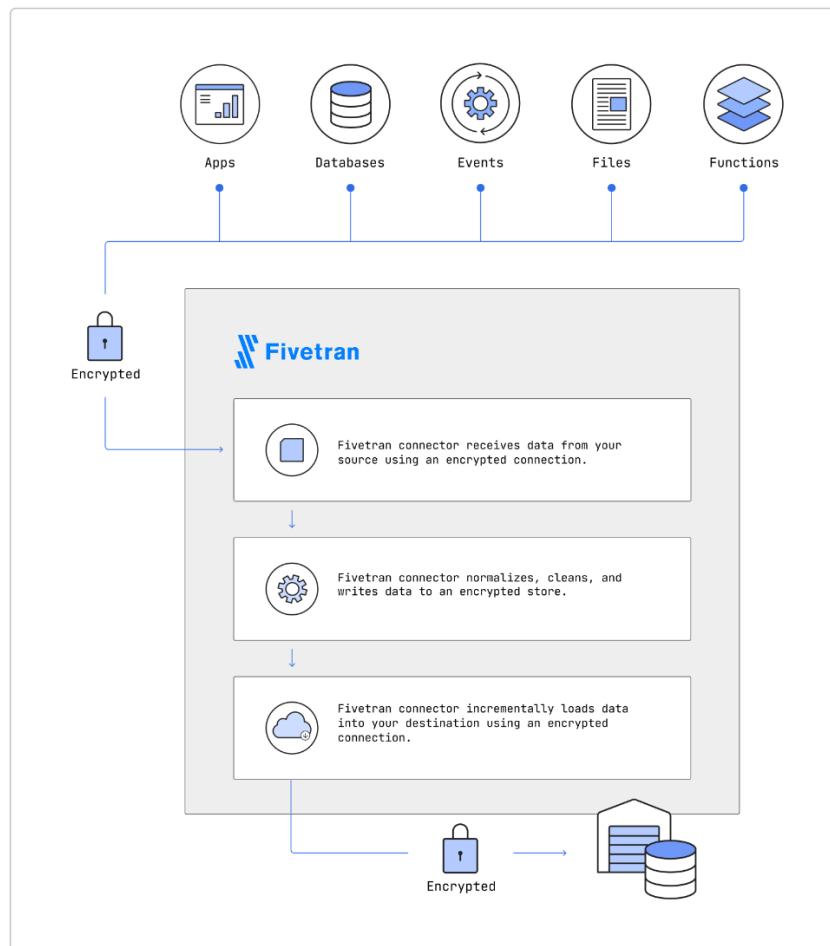


Ilustración 1: Arquitectura de Fivetran

Referencias

[Documentación](#)

[Guía Aprendizaje](#)

[Comunidad Fivetran](#)

[Overview de Fivetran](#)

Acceso a la herramienta

Hay que crear una cuenta en su página web (<https://fivetran.com/signup>) , que tiene una versión trial de 14 días para iniciar en la plataforma; además podemos ver una tabla de diferentes planes de pago que tienen, ya que se puede elegir un plan concreto (existe uno gratuito para volúmenes bajos de datos) o elegir por uso.

PRECIOS

<p>Free For individuals automating ELT for small volumes of data.</p> <p>Start for free</p> <p>✓ Access Standard Plan features, free up to 500,000 monthly active rows (MAR) ✓ Commitment free: No credit card required</p> <p>Learn more →</p>	<p>Starter Achieve modern ELT for your application and file sources.</p> <p>Start for free</p> <p>✓ 10 users ✓ 1-hour syncs ✓ 400+ fully-managed connectors ✓ Automatic schema migrations ✓ Integration for dbt Core ✓ Role-based access control ⓘ</p>	<p>Standard Centralize product and transactional data from your databases.</p> <p>Start for free</p> <p>✓ Everything in our Starter Plan, plus: ✓ Unlimited users ✓ 15-min syncs ✓ Database connectors ✓ Access to Fivetran's REST API</p>	<p>Enterprise Empower everyone with fast, secure, reliable data from all of your sources.</p> <p>Start for free</p> <p>✓ Everything in Standard, plus: ✓ Enterprise database connectors ⓘ ✓ 1-minute syncs ✓ Granular roles & support for teams ⓘ ✓ Advanced data governance ⓘ ✓ Advanced security and data residency options ✓ Priority support</p>
--	---	---	---

Ilustración 2: Planes de Fivetran

2. OVERVIEW DE LA HERRAMIENTA

1. Automatización Integral

Fivetran automatiza el proceso de extracción, transformación y carga (ETL) de datos, eliminando la necesidad de intervenciones manuales.

2. Conectores Preconstruidos

Ofrece más de 400 conectores preconfigurados para diversas fuentes de datos, facilitando la conexión con una amplia variedad de aplicaciones, bases de datos y herramientas.

3. Sin Necesidad de Código

Proporciona una interfaz sin código que permite a los usuarios configurar y gestionar integraciones sin conocimientos avanzados de programación.

4. Gestión de Cambios Automatizada

Maneja automáticamente las derivas de esquemas, actualizaciones y normalización de datos, asegurando la coherencia y precisión de la información.

5. Captura de Cambios Eficiente

Utiliza técnicas de captura de cambios de alto rendimiento para realizar movimientos de datos eficientes y de bajo impacto, reduciendo el tiempo y los recursos necesarios.

6. Seguridad y Escalabilidad

Prioriza la seguridad de los datos y garantiza la escalabilidad para adaptarse a las necesidades crecientes de las empresas.

7. Soporte para la Nube

Diseñado para entornos basados en la nube, facilita la integración de datos desde y hacia plataformas en la nube como AWS, Google Cloud, Azure y otros.

8. Análisis Centrado en Datos

Permite a las organizaciones concentrarse en el análisis y la toma de decisiones al proporcionar datos precisos y actualizados de manera centralizada.

3. CASO PRÁCTICO

Vamos a explorar un ejemplo práctico para evaluar el desempeño y las características de las diversas transformaciones proporcionadas por Fivetran. En este caso, nos enfocaremos en crear un datamart sencillo basándose en la base de datos de Adventure Works.

Para que se inicie la prueba de 14 días, tenemos que elegir el conector que queremos usar, el destino de los datos y hacer una prueba de que funcione todo correctamente.

La primera vez que accedamos, tendremos una interfaz de esta manera:

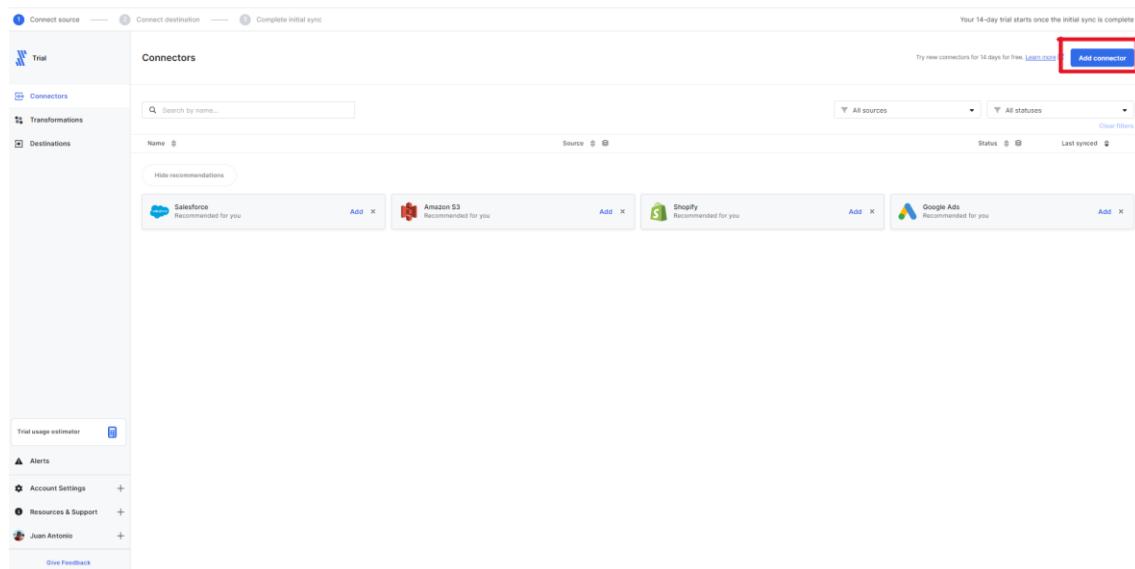


Ilustración 3: Interfaz principal de inicio

OPCIONES DESDE LA INTERFAZ

1. DESTINATIONS
2. CONNECTORS
3. TRANSFORMATIONS
 - A. Quickstart data models. - Nativa
 - B. Dbt Core Integration

1. DESTINATIONS

Pulsaremos donde está el recuadro azul para empezar la configuración del **DESTINO**

Haremos este ejemplo con un destino a una base de datos de Microsoft SQL Server

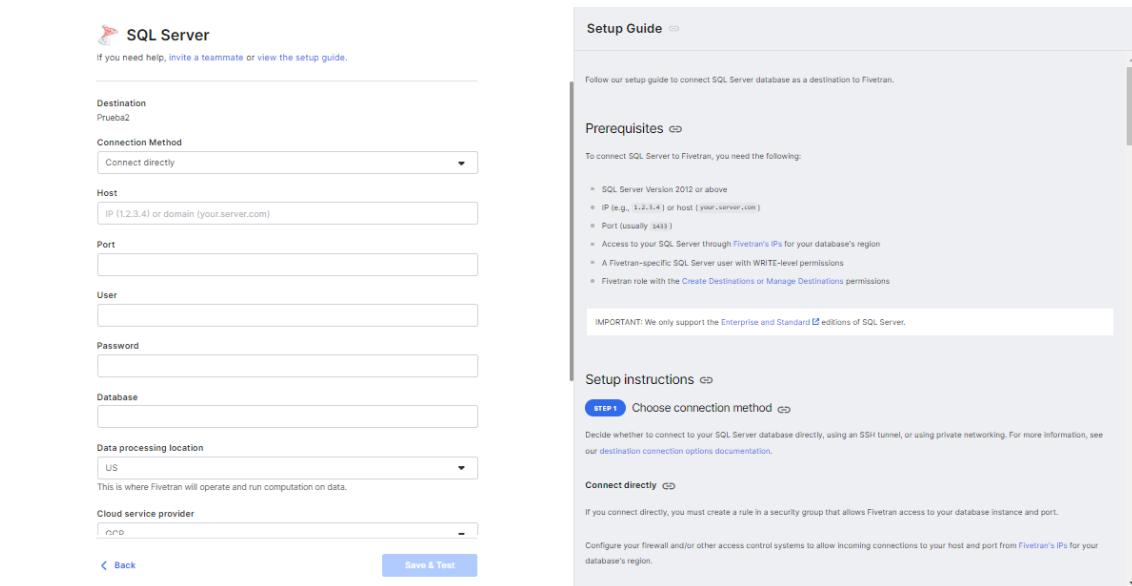


Ilustración 4: Configuración del Connector MY SQL SERVER

Tras configurar los parámetros de conexión de la base de datos destino, si pasa todas las comprobaciones tendremos en el apartado “Destinations”.

Destinations			
1 Connected			
Add destination			
<input type="text" value="Search by name..."/>			
Name	Type	Status	Admins
Prueba_SQLServer	SQL Server	Connected	Juan Antonio Perez Cifuentes

Ilustración 5: Lista de Destinos Configurados

2. CONNECTORS

Tras esto, tendremos que configurar en el apartado “Connectors”, pulsaremos en el recuadro azul para empezar la configuración, nosotros hemos elegido también SQL Server como base de datos de origen y tras hacer la configuración tendremos ya configurado lo imprescindible.

[← Back](#)

Select your destination

4 destinations		Sort by: Default ▾
	Destino1	
	Prueba_SQLServer	
	PruebaLocal	

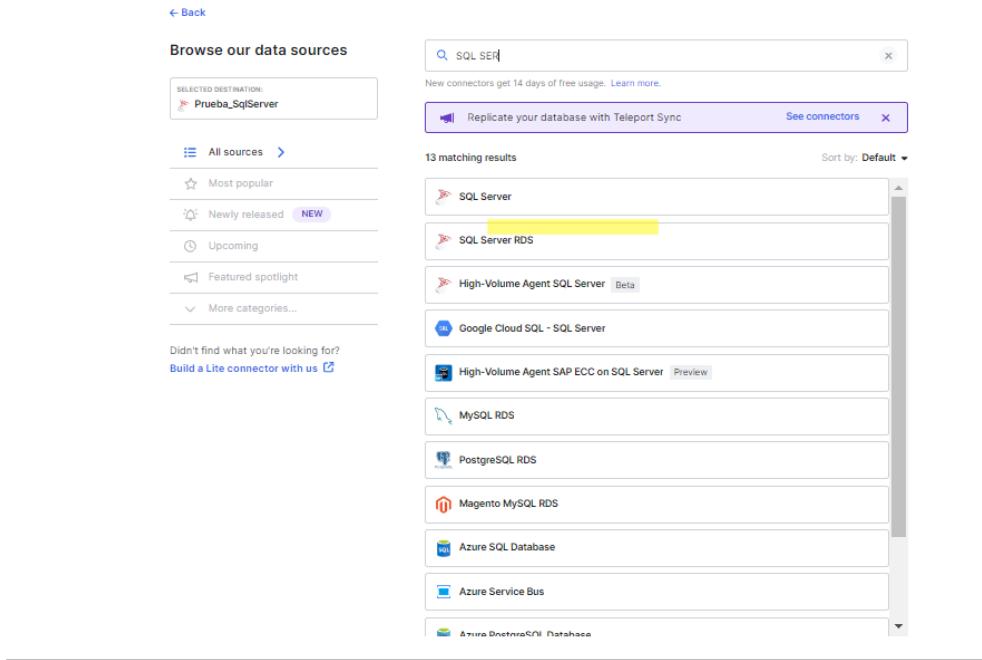


Ilustración 6: Configuración de la Base de Datos Destino

Se hará de la misma manera que se hizo en el apartado “Destinations” la configuración de la base de datos “Destino” y se le asignará un nombre de schema donde el cuál se meterán todas los schemas/bases de datos que queramos orquestar.

Tras esto tendremos que hacer la sincronización inicial para que realice la orquestación de los datos de la base de datos origen a la base de datos destino; tras esto podremos configurar cada cuánto tiempo se realice este proceso.

Ilustración 7: Lista de Conectores en Estado “Completado”

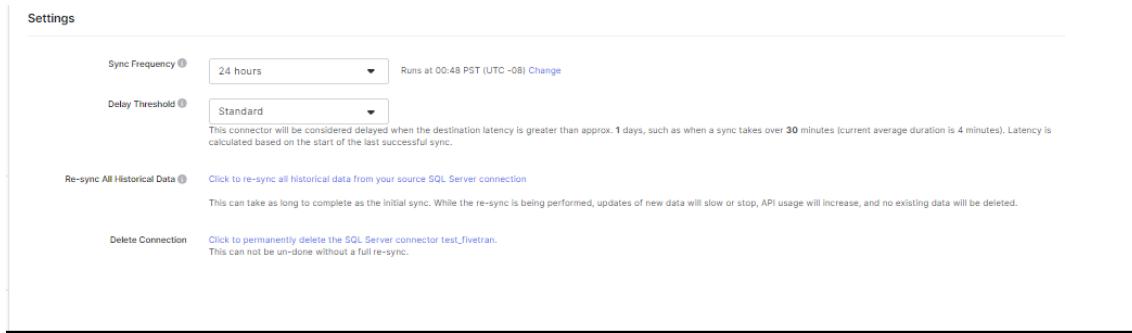


Ilustración 8: Configuración del tiempo de sincronización

Como podemos ver en la imagen, tenemos todos los esquemas y tablas sobre Adventure Works y si miramos en el Dbeaver la base de datos destino podemos ver las tablas que se han creado.

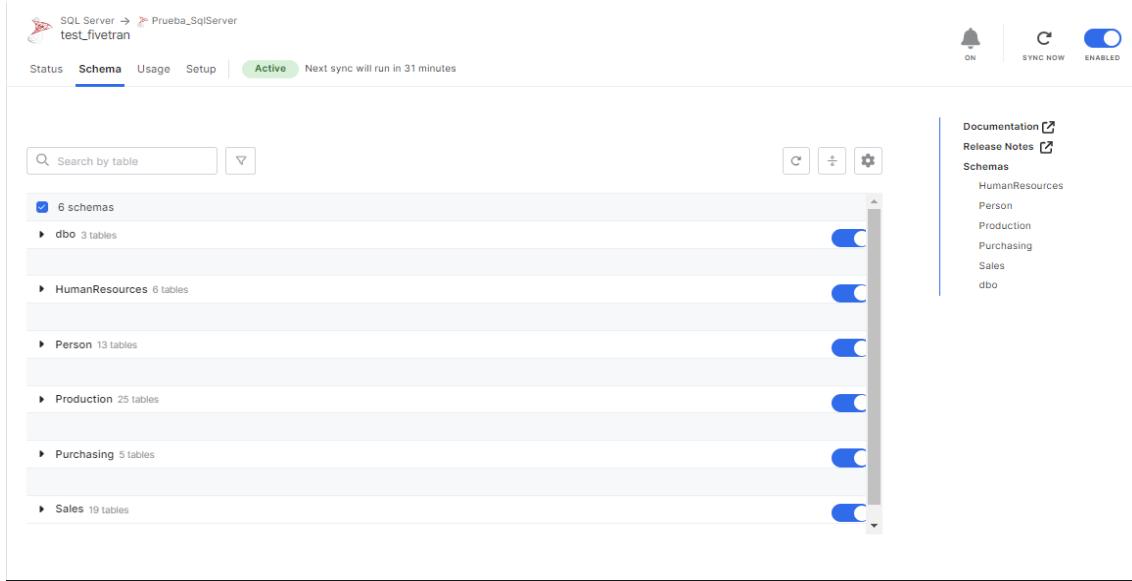


Ilustración 9: Lista de los esquemas y tablas copiadas de la base de origen a la base destino.

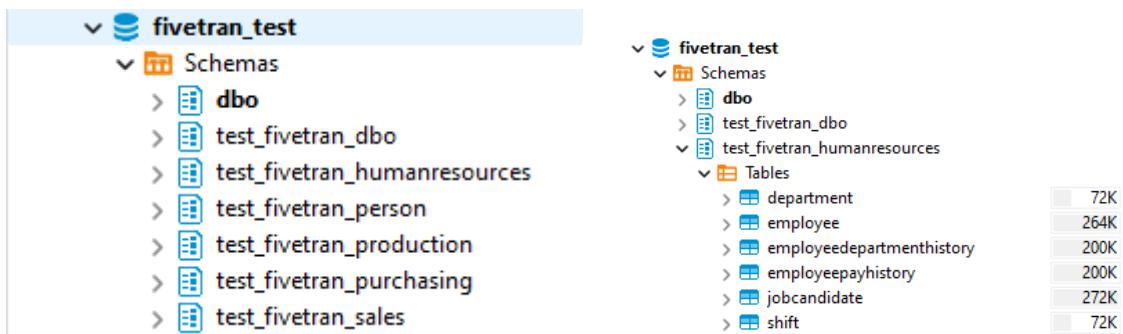


Ilustración 10: Listado de los schemas y tablas desde el Dbeaver.

3. TRANSFORMATIONS

El siguiente paso es entrar en “Transformations”.

Aquí la herramienta nos ofrece 2 opciones:

Select transformation type

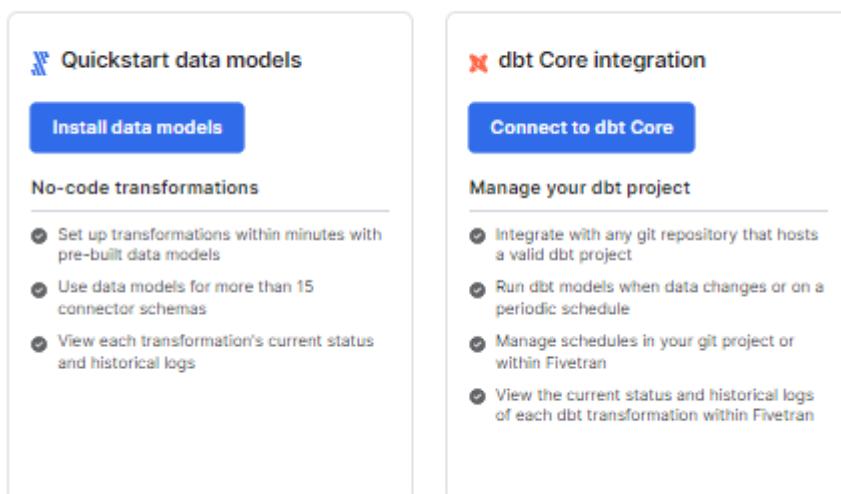


Ilustración 11: Opciones de Transformations

1. Quickstart data models.

Fivetran tiene varios conectores ya predeterminados, los cuales basta con introducir las credenciales de ese conector te realiza las transformaciones necesarias y no tienes que utilizar ningún código, no es una lista muy extensa pero a cierto sector le automatiza y le ahorra bastante tiempo.

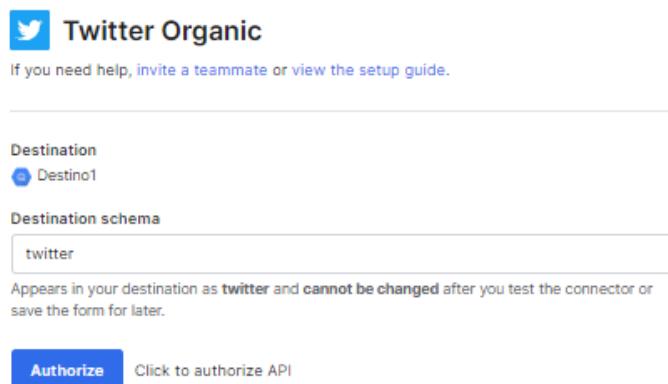
CONECTORES

- Amplitude
- Facebook Ads
- Fivetran Platform Connector
- Google Ads
- Greenhouse
- Twitter
- Stripe

El inconveniente sobre este tipo de transformación es que las únicas bases de datos admitidas son:

- BigQuery
- Databricks
- PostgreSQL
- Redshift
- Snowflake

A. TWITTER



The screenshot shows the Fivetran interface for setting up a Twitter Organic connector. At the top, there's a Twitter logo and the text "Twitter Organic". Below that, a link to "If you need help, invite a teammate or view the setup guide." The main configuration area has two sections: "Destination" and "Destination schema". Under "Destination", there's a dropdown menu with "Destino1" selected. Under "Destination schema", there's a text input field containing "twitter". A note below the schema field states: "Appears in your destination as **twitter** and **cannot be changed** after you test the connector or save the form for later." At the bottom, there's a blue "Authorize" button with the sub-instruction "Click to authorize API".

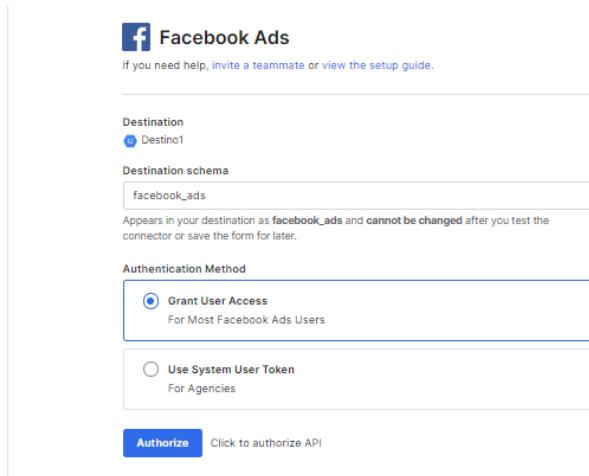


Ilustración 12: Configuración del Connector Twitter

Este conector directamente ya tiene las [tablas](#) que se van a crear y simplemente lo único necesario para que funcione es permitir acceder a tu propia cuenta de Twitter.

B. FACEBOOK ADS

Este conector esta enfocado para estudiar las gráficas que podamos tener en nuestro perfil de Facebook y la configuración es similar a la de Twitter ; únicamente teniendo que autorizar que Fivetran se conecte a nuestra cuenta de Facebook y diciendo un nombre de schema donde irán todas las tablas.



Tablas de Facebook

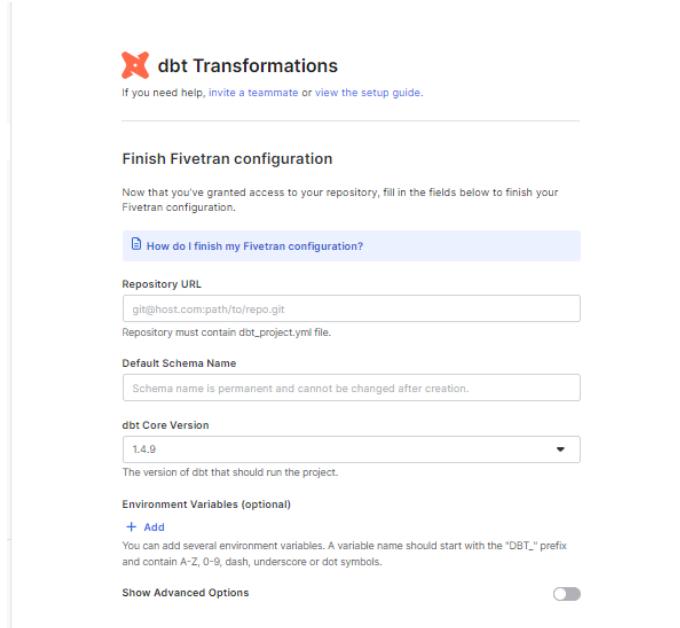
Ilustración 13: Configuración del Connector Facebook Ads

2. **Dbt Core Integration**

Esta opción es la más flexible ya que nos permite crear código y tener más libertad en la creación de nuestras transformaciones.

Esta consiste tener nuestro proyecto DBT en un repositorio GitHub el cuál vincularemos con Fivetran para poder ejecutar las transformaciones creadas.

Habrá que configurar en nuestro propio repositorio la clave SSH que nos otorgan ellos para poder darles acceso y llenar los datos necesarios para poder hacer la conexión.



The screenshot shows the 'dbt Transformations' configuration interface. At the top, there's a logo and the text 'dbt Transformations'. Below that, a message says 'If you need help, invite a teammate or view the setup guide.' The main section is titled 'Finish Fivetran configuration'. It says 'Now that you've granted access to your repository, fill in the fields below to finish your Fivetran configuration.' There's a 'How do I finish my Fivetran configuration?' link. The configuration fields include:

- Repository URL:** git@host.com:path/to/repo.git
- Default Schema Name:** Schema name is permanent and cannot be changed after creation.
- dbt Core Version:** 1.4.9
- Environment Variables (optional):** A section with a '+ Add' button and a note: 'You can add several environment variables. A variable name should start with the "DBT_" prefix and contain A-Z, 0-9, dash, underscore or dot symbols.'
- Show Advanced Options:** A toggle switch.

Ilustración 14: Configuración para la vinculación a Git

Add new SSH Key

Title

Key type Authentication Key

Key

Begins with 'ssh-rsa' 'ecdsa-sha2-nistp256' 'ecdsa-sha2-nistp384' 'ecdsa-sha2-nistp521' 'ssh-ed25519' 'sk-ecdsa-sha2-nistp256@openssh.com' or 'sk-ssh-ed25519@openssh.com'

Ilustración 15: Añadir la clave SSH que nos proporciona Fivetran

Información sobre DBT para Microsoft SQL Server

Aquí están enlaces que explican un poco más sobre la herramienta que se utiliza para crear las dimensiones y la tabla de hecho, además este ejemplo se ejecutó con una base de datos la cual el conector fue creado por la propia comunidad.

[DBT](#)

[Conectores Comunidad DBT](#)

[ODBC Driver 18 for SQL Server](#) “Necesario para la creación en DBT de los Jobs”

Se muestran una captura de un ejemplo de la creación de una Dimensión y como quedaría unido todo gracias al comando.

- **dbt docs serve**

```
fivetran > models > d_scrapreason.sql
1  {{config(materialized='table')}}
2
3  WITH datosscrapped AS (
4      SELECT
5          sr."scrapreasonid",
6          sr."name"
7      FROM {{ source('Produccion','scrapreason') }} as sr
8
9  )
10
11  SELECT * FROM datosscrapped
```

Ilustración 16: Dimensión datosScrapped

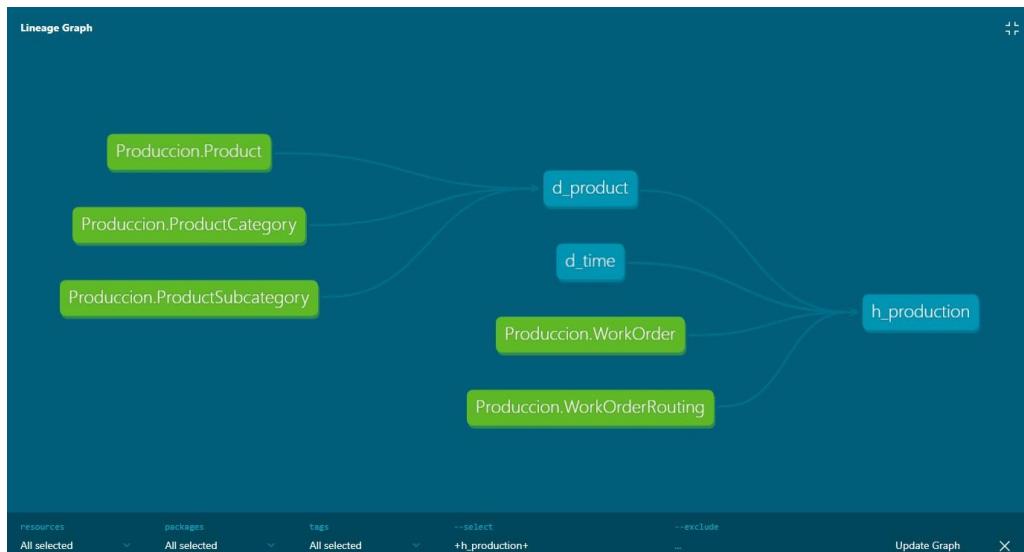


Ilustración 17: Linaje del DBC

Tras la configuración correcta ya podremos hacer transformaciones desde Fivetran.

Seleccionaremos todas las dimensiones y la tabla de hecho que hemos creado gracias a DBT.

Custom transformation

Choose a model and create a schedule for Fivetran to run that model.

Git repo connected: git@github.com:JuananPerez/fivetran.git

Choose model

Model

- d_tiempo [Output model]
- stg_dbt_deployments [Output model]
- stg_dbt_model_deployments [Output model]
- stg_dbt_audit_log
- h_production [Output model]

Destination: Prueba_SQLServer

Transformations BETA

Connected | Git sync completed 3 minutes ago

All transformations are free in Fivetran. [Learn more](#) [Settings](#) [Add transformation](#)

Name	Type	Connectors	Run frequency	Status	Last run
d_scrapreason	dbt	None	Synced with connectors	Pending	-
d_tiempo	dbt	None	Daily at 9:00 AM	Pending	-
h_production	dbt	None	Synced with connectors	Pending	-
d_producto	dbt	None	Synced with connectors	Pending	-

Ilustración 18: Conjunto de transformaciones para ejecutar

Tras esto, ejecutamos cada transformación y podremos ver que se ha creado el datawarehouse en nuestra base de datos.

fivetran_test

Schemas

datawarehouse

Tables

- d_producto 200K
- d_scrapreason 200K
- h_production 264K

Ilustración 19: Tablas tras la ejecución de las transformaciones en Fivetran

4. VENTAJAS/DESVENTAJAS

1. **Automatización:** Fivetran automatiza gran parte del proceso de la orquestación de datos, desde la extracción de datos hasta la carga en el destino. Esto reduce la carga de trabajo manual y minimiza los errores.
2. **Mantenimiento Automático:** La plataforma se encarga del mantenimiento y la actualización de los conectores, lo que libera a los usuarios de la tarea de gestionar y mantener las integraciones a lo largo del tiempo.
3. **Escalabilidad:** Puede escalar fácilmente para manejar grandes volúmenes de datos a medida que las necesidades de la empresa crecen. Esto es fundamental para empresas en crecimiento que experimentan un aumento en la cantidad de datos.
4. **Seguridad:** Fivetran implementa prácticas de seguridad robustas para proteger los datos durante el proceso de integración. Esto incluye la encriptación de datos en tránsito y en reposo, así como otras medidas de seguridad.
5. **Monitoreo y Alertas:** Proporciona herramientas de monitoreo que permiten a los usuarios supervisar el estado de las integraciones y recibir alertas en caso de problemas.

DESVENTAJAS

1. Flexibilidad y Personalización:

Fivetran, al ser una plataforma sin código, puede carecer de la flexibilidad y personalización que algunos proyectos más complejos requieren. PDI, al ser una herramienta de código abierto y con capacidad para scripts, permite una mayor adaptabilidad para casos específicos.

2. Control Detallado del Proceso ETL:

Para proyectos que requieren un control más detallado del proceso ETL, PDI puede ser más adecuado. Fivetran automatiza gran parte del proceso, pero no permite usar código para cambiar una transformación.

3. Modelo de Precios:

Aunque Fivetran tiene un modelo de precios relativamente simple, podría no ser la opción más rentable para todas las situaciones. En comparación, PDI al ser de código abierto puede ser más económico en términos de licencias ya que se puede usar la versión gratuita con escasas limitaciones y especialmente para organizaciones que prefieren gestionar su infraestructura.

4. Capacidades de Procesamiento en el Lugar (On-Premise):

Fivetran está optimizado para entornos basados en la nube, y puede no ser la mejor opción para aquellos que requieren una solución exclusivamente en las instalaciones (on-premise). PDI, por otro lado, ofrece más flexibilidad en cuanto a dónde se ejecutan los procesos ETL.

5. Requerimientos de Conectividad:

Fivetran depende de conectores preconstruidos, lo que facilita la conexión con una amplia variedad de fuentes de datos. Sin embargo, para casos extremadamente específicos o tecnologías menos comunes, PDI podría ser más adaptable a través de configuraciones manuales y personalizadas.

6. Complejidad para Usuarios Técnicos Avanzados:

Para usuarios técnicos avanzados que prefieren un control total sobre el código y los scripts, la simplicidad de Fivetran podría percibirse como una limitación. PDI, con su capacidad de scripting y desarrollo más detallado, puede ser más atractivo en estos casos.

5. CONCLUSIÓN

En resumen, Fivetran destaca por su simplicidad, automatización total y enfoque en la nube, lo que puede ser beneficioso en situaciones donde se busque una implementación rápida y eficiente sin una carga técnica significativa. PDI, por otro lado, ofrece una mayor flexibilidad y control para aquellos con requisitos más específicos y habilidades técnicas avanzadas. La elección entre ambos dependerá de las necesidades y el contexto específico de cada proyecto.

1. Facilidad de uso:

- Fivetran generalmente se destaca por su enfoque en la simplicidad y la facilidad de uso. Ofrece conectores preconstruidos para diversas fuentes de datos, lo que facilita la configuración de la integración.
- PDI es más flexible y puede ser personalizado para satisfacer necesidades específicas, pero esto también puede hacer que la curva de aprendizaje sea más pronunciada.

2. Conectividad y Fuentes de Datos:

- Fivetran proporciona conectores listos para usar para una amplia variedad de fuentes de datos en la nube y en las instalaciones.
- PDI es más versátil en términos de conectividad y puede integrarse con una amplia gama de fuentes de datos, pero podría requerir más configuración manual.

3. Escalabilidad:

- Fivetran está diseñado para ser escalable y gestionado fácilmente en entornos en la nube.
- PDI puede ser escalado según las necesidades, pero la implementación y gestión pueden requerir más esfuerzo.

4. Costos:

- Los modelos de precios de Fivetran suelen ser basados en el volumen de datos y el número de conectores utilizados.
- PDI, al ser de código abierto, puede ser más económico en términos de licencias, pero los costos asociados con la implementación y el mantenimiento pueden variar.

Paper creado por [stratebi](#) (AKA Data Ninjas)