How to choose the most cost-effective data pipeline for your business

Your ELT solution should be engineered to optimize your stack and make your team more efficient

Fivetran

Table of contents

I. Introduction: Leverage ELT to increase the efficiency of your data team and stack 3

II. Building data pipelines: Is DIY cost effective? 5

III. Maintaining data pipelines: If you buy, will you still have to fix connectors? 10

IV. Moving data: Key attributes of a highly efficient data pipeline 13

V. Transforming data: How an ELT platform can accelerate analytics 22

VI. Conclusion: Test ELT efficiency before you commit 28

INTRODUCTION

I.

Leverage ELT to increase the efficiency of your data team and stack

In the current economic climate, **your data team is probably being asked to do more with less.**

Data has become a tremendous competitive advantage, and it's especially valuable during uncertain economic times, helping organizations survive and even thrive through adversity. Your team is likely being asked to handle more data from more sources while reducing time to insight — without additional budget or headcount.

How do you resolve that dilemma?

In this ebook, we'll take a close look at how a data movement platform can improve your analytics outcomes while making your data team — and stack — more efficient. We'll start with the pipeline building process, move on to pipeline maintenance, and then consider data movement and data transformation.

MORE SPECIFICALLY, WE'LL ANSWER THESE QUESTIONS:

1. Will DIY data pipelines limit the efficiency of your data team or your organization's business agility? Is building them yourself worth the investment?
2. How much pipeline maintenance would an ELT tool or platform require your team to do?
3. Does it use automation to reduce or eliminate maintenance?
4. Would its approach to syncing data increase your data warehouse costs? Does it perform incremental syncs and use idempotence to prevent data duplication when syncs fail?
5. Does it facilitate data modeling with thoughtful ERDs and prebuilt data models?
6. How does it approach data transformation? Does it offer integrated scheduling for connector syncs and transformation runs in a single UI?

At Fivetran, we've spent many years working through these challenges and developing capabilities that make data teams more efficient and effective. Ultimately, these kinds of capabilities determine whether a data movement tool or platform is cost-effective — whether, in other words, it's going to make life a lot easier for your data team and reduce your data stack costs.

By the end of the ebook, you'll understand the key determinants of data movement cost-effectiveness, so that whatever solution you choose, you'll know exactly which tradeoffs you're making and why.

II.

Is DIY cost effective?

Many organizations build their own data pipelines, and if you have the in-house expertise, that can seem like a cost-effective solution. DIY pipelines may also give you more control over the flow of your data.

There are significant drawbacks, however — data engineering opportunity cost chief among them. The pipelines will have to be built by your own engineers, which means they can't be simultaneously working on higher-value projects.

The first question you should ask about DIY is: **How many data engineering hours will we have to devote to the project?** A second key consideration is the impact of DIY on scalability and business agility — how rapidly you can add new data sources and train new engineers.

Data engineering hours it takes to DIY a data pipeline

THE BUILD PROCESS

The time required to complete data pipeline projects varies widely. Factors like API complexity and the number of individual data connectors comprising the pipeline[1] can add or subtract quite a few hours. Even modest pipeline projects, however, tend to take weeks or months to complete.

$520,000 per year

Average cost of building and maintaining data pipelines for a business with 12 engineers

SOURCE: WAKEFIELD RESEARCH

[1] We're defining a data pipeline as the collection of individual data connectors required to centralize an organization's data. A pipeline could consist of **NetSuite**, **Marketo** and **Salesforce** connectors, for example — or many more.

DIY data connectors tend to be significantly less advanced than Fivetran data connectors. **If you attempt a more automated and robust connector, it could easily take six months of engineering time.**

The stages of the build process would look like this:

1. Explore and document source data
2. Orchestration and automated scheduling
3. Incremental updates
4. Idempotence (recovery from failure)
5. Lossless replication
6. Logic and process isolation
7. Distributed architecture
8. Role-based access control
9. Security and regulatory compliance
10. Automated schema migration

Canva

Greg Roodt, Head of Data Platforms at **Canva**, says that before the company started using Fivetran, three of its engineers spent three months creating, testing and deploying custom connectors for **Braze**, **AppsFlyer** and **Apple App Store**.

> They're excellent engineers, but they had to slog through confusing and outdated API documentation, set up the infrastructure in AWS, create a monitoring mechanism and make sure it all worked."

jetBlue

Ashley Van Name, General Manager of Data Engineering at **JetBlue,** says that:

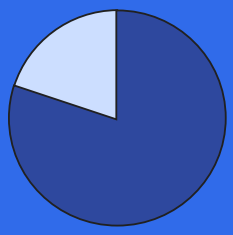
> Before Fivetran, it regularly took engineers weeks, if not months, to fully build, test and deploy data pipelines."

THE MAINTENANCE PROCESS

Because they're not automated and lack self-healing capabilities, DIY pipelines tend to require substantial maintenance after construction.

Stationary startup **Papier** CTO Joe Robertson also spent at least a full day a week fixing ETL issues caused by DIY pipelines. Now he never thinks about it: "It's been fantastic to just let Fivetran run."

80% of data and analytics leaders rebuild data connectors after deploying.



SOURCE: WAKEFIELD RESEARCH

We built a **TCO calculator for DIY pipelines**, which tells a similar story. Sanity-checked by Fivetran data engineers, it shows that **building and maintaining three DIY connectors, for example, would cost $75,000 and require six months of engineering time — in the first year of operation.**

Keep in mind that the average business currently **relies on somewhere between five and 20 data sources** — and the average organization **relies on 400**.

datacamp

> You could, in theory, build pipelines to get data into your data lake or warehouse yourself, before pointing out this requires, at a minimum, reading the docs, learning how the API works, writing the code and moving the data to your destination."

Gaëtan Podevijn, data engineer at DataCamp

EMERY SAPP & SONS

> I spent at least a day a week on pipeline management. Without Fivetran, I'd have to hire a dedicated engineer to maintain our current stack. I'm saving so much time that I actually shudder thinking about the past."

Clayton Hicklin, Director of Technology at construction firm Emery Sapp & Sons

The steep opportunity cost of DIY data pipelines

Your data engineers are accounted for in your current budget, but the value you lose when you allocate their talent to data pipeline work is almost certainly the most powerful argument against building data pipelines yourself.

In the Wakefield Research survey, 70 percent of data leaders said their engineers would add more value to the business if they could focus on strategic projects instead of pipeline construction and maintenance, including:

- Working on advanced data models and advanced analytics
- Building internal applications to optimize business functions
- Researching new data tools
- Training coworkers on data practices and technologies

The scalability problem: DIY makes it difficult to add data sources and engineers

Scalable data pipelines are vital for business growth — and as your company grows, you'll need pipelines that can handle more data sources and higher data volumes, as well as complex performance requirements.

It's not realistic to expect your engineers to build and maintain a connector for every new data source, let alone add sophisticated new features and automations.

"We're an ecommerce company, but we don't have an army of engineers. Not having to pull in or manage data frees up engineering resources for more critical projects that are our core business."

Chao Li, data scientist at mattress company Sleeping Duck

WESTWING

"Fivetran saves us 40 hours of engineering time a week, so we can focus more on the strategic vision."

Christian Manrich, Business Intelligence Engineering Manager at home furnishing company Westwing

It's also much harder to onboard new engineers if you have DIY pipelines. They will have to learn the unique API documentation for every source, as well as understand how the entire pipeline is structured — so if, for example, a source API changes, they know how to alter any subsequent transformations to keep data flowing.

The penalty of using open-source software

You do have the option of building pipelines using an open-source connector development kit (CDK). Open-source technologies are free to use, designed for ease of integration within the data platform, and offer an alternative to manual coding, which can speed up the build process considerably. You will run into a host of inefficiencies that can slow your team down, however, and you may incur hidden costs as well. Here are the main open-source shortcomings to be aware of.

BUGS AND INADEQUATE DATA

Because open-source offerings generally aren't subjected to rigorous QA testing, they tend to be buggy. They also sync far fewer tables than more advanced connectors — the data you need may or may not be available.

HIDDEN INFRASTRUCTURE AND COMPUTE COSTS

Generally, you'll need to host open-source connectors on your own infrastructure, so some costs will be associated with their use. They can also increase monthly compute costs by as much as a factor of three compared to Fivetran, according to several current Fivetran customers.

LACK OF AUTOMATION AND LOW DATA QUALITY

Open-source connectors do not support automated schema migration for columns or data types — but these changes occur frequently, making it extremely difficult to ensure data integrity, with data in the source rarely matching data in the destination.

Schema changes often require a full resync, which can bring down reporting for days.

LACK OF SUPPORT

Since open-source software is free, it is typically supported outside of a user community. It cannot offer enterprise-grade support in the form of uptime, reliability and SLAs.

III.

If you buy, will you still have to fix connectors?

As you think about the right data movement solution for your team, **pipeline reliability — in particular how a pipeline handles schema and API changes — is a critical factor to consider.**

Pipeline breaks due to data source changes can take hours or days to repair, as engineers have to identify what has broken, fix it and understand any downstream effects the break may have caused. Even if a pipeline doesn't completely break, data teams have to make code and configuration updates to capture the new changes and redeploy the pipeline.

Delayed data caused by pipeline issues results in key metrics becoming outdated and untrustworthy, compromising decision-making or paralyzing the decision-making process alltogether. Consequent missed revenue opportunities can be extremely damaging.

As one data leader at an ecommerce company **recently observed**, "It's worse to have bad data on my company's website than to not have a website at all."

Detecting and handling data source changes

Reliable data replication is fundamentally about faithfully preserving the original values of data and ensuring its smooth passage from source to destination, regardless of schema and API changes. Yet many types of data movement tools — DIY and open-source generally, as well as some existing ELT solutions — fail to adequately address the challenge of building and maintaining a robust and reliable data pipeline.

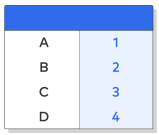
A maximally efficient data movement platform like Fivetran will protect your data team's time by detecting and handling source changes automatically.

SCHEMA CHANGES

Open-source connectors do not support automated schema migration for columns or data types. Even a small schema change can have disproportionately large downstream effects — it may take many hours to change the affected transformations, manually edit the destination schema so new data can flow through, etc. In the meantime, dashboards that functional teams rely on may break and incomplete or inaccurate data may appear in reports to executives or other stakeholders.

Fivetran eliminates these issues, and the work required to address them, by **automatically detecting and responding to source schema changes** — no intervention on your part is required.

More specifically, we incrementally and comprehensively propagate schema changes from the data source to your data destination. Here's a quick rundown of how Fivetran handles common schema changes:

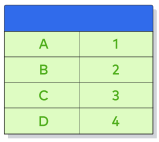


New column

If your source adds a new column, we detect the change and add the same column in our warehouse, backfilling the data if applicable

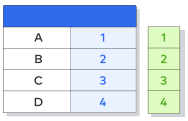
Deleted column

If a column is removed, we won't delete it outright from your data warehouse but "soft-delete" it and mark future records NULL so that the old data remains as a historical point of reference.



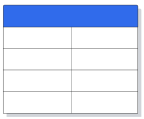
New table

If your source adds a new table, we will begin syncing it like any other.



Type changes

If a column changes data types in the source, we create a new column with a temporary name, copy all the values into it and cast them as the most specific data type that losslessly accepts both the old and new data, drop the old column, and then rename the new column with the name of the dropped column.



Deleted table

If a table is deleted in your source, we will leave it alone in your warehouse.

API CHANGES

When API data models change, the Fivetran engineering team — over 320 engineers across the globe — rapidly addresses those changes in the connector code, sparing your team the time cost of broken pipelines and the business impact of inaccurate or outdated metrics.

Monitoring and patching your pipeline environment

It takes time to monitor the multiple pieces of infrastructure that run your pipelines, and additional time to patch your environment as needed. Whether those responsibilities fall entirely on your data team or not, your data movement provider will — ideally — take them completely off your hands, ensuring that you avoid worst-case scenarios involving outdated infrastructure components or missed security patches.

That's what Fivetran does, managing and supporting the entire data movement process 24/7/365, with a guaranteed uptime of 99.9 percent.

Your team won't have to worry about operating system or other upgrades, or security patches — we rigorously maintain our security credentials (GDPR, CCPA, SOC 2, ISO, HIPAA,PCI, HIPAA), so you don't have to spend time on security audits.

blend

Cloud banking infrastructure provider Blend relied entirely on DIY pipelines to centralize data from multiple sources — and those pipelines were extremely vulnerable to source changes.

> "We were hitting enough data where it took a very long time to load into our system and push back out, or our pipelines were just not working any more. **One of the things we love about Fivetran is the schema drift handling in the database, with no reliance on engineering.** I don't have to file a ticket just to get a column added."

William Tsu, Head of Customer Success Operations at Blend

IV.

Key attributes of a highly efficient data pipeline

Eliminating the engineering burden of building and maintaining data pipelines provides a massive boost to the efficiency and effectiveness of data teams, but ELT platforms can offer a set of equally impactful capabilities when it comes to moving and managing your data.

As you and your team consider scaling data integration, it's important to understand how an ELT tool or platform can save you time and make your stack more efficient in five key areas:

Data normalization

Idempotence (recovery from failure)

Incremental syncing

Granular data blocking

Programmatic pipeline management

The right engineering approach in all these areas will have a powerful cumulative effect on the performance and efficiency of data teams and stacks. Combined with liberation from pipeline maintenance, they'll also have a nontrivial quality-of-life impact — data teams are just happier when they can focus on interesting, high-value work and forget about manual engineering chores or suddenly soaring warehouse costs.

Let's take a look at best-practice data engineering across each of these capabilities.

Data normalization

Semi-structured data from a SaaS source like Marketo is not immediately usable for SQL-based data analysis — but what's the best way to normalize it into a standard format and create a schema that is maximally useful to data engineers and analysts? ELT tools can do this well, or poorly, or not at all, and the difference will have a huge impact on the efficiency of data modeling and data warehouse performance.

Designing a schema that delivers accurate, comprehensive data is quite painstaking and time-consuming — and this process is partially repeated each time the underlying data model of an application changes.

Our systematized workflow for designing schemas is the result of thousands of hours of ongoing work and learning. We take pride in sparing your data team this tedium and offering something that works with the push of a button.

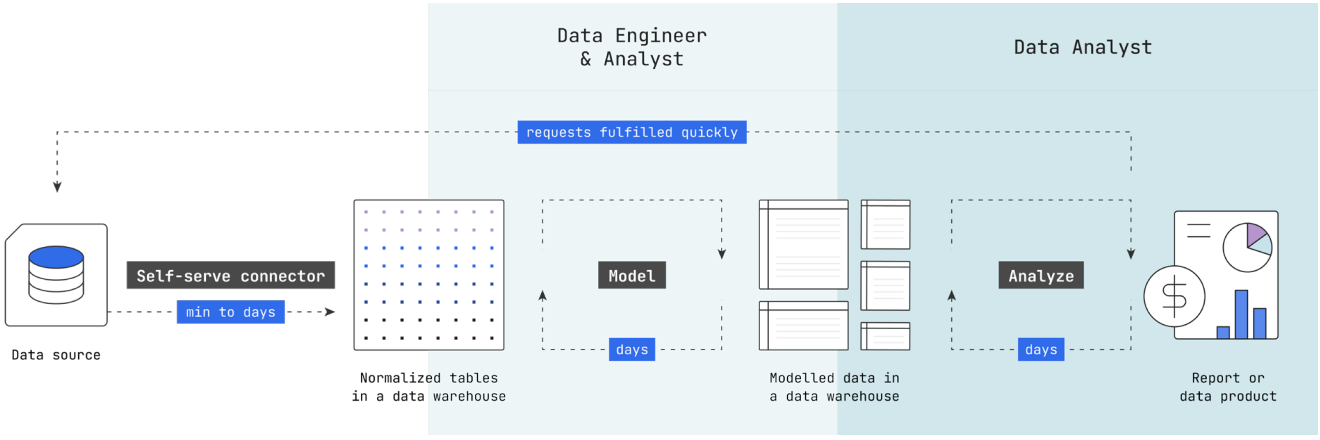
BENEFITS OF WELL-DESIGNED SCHEMAS

A **well-designed normalized schema** provides nearly everything someone needs to know to work with a data set:

- The tables represent business objects
- The columns represent attributes of those objects
- Well-named tables and columns are self-describing
- The primary and foreign key constraints reveal the relationships between the business objects

Fivetran schemas are as close to third normal form (3NF) as possible.[2] For SaaS API connectors, we've designed the schemas to represent the underlying relational data model in that source.[3]

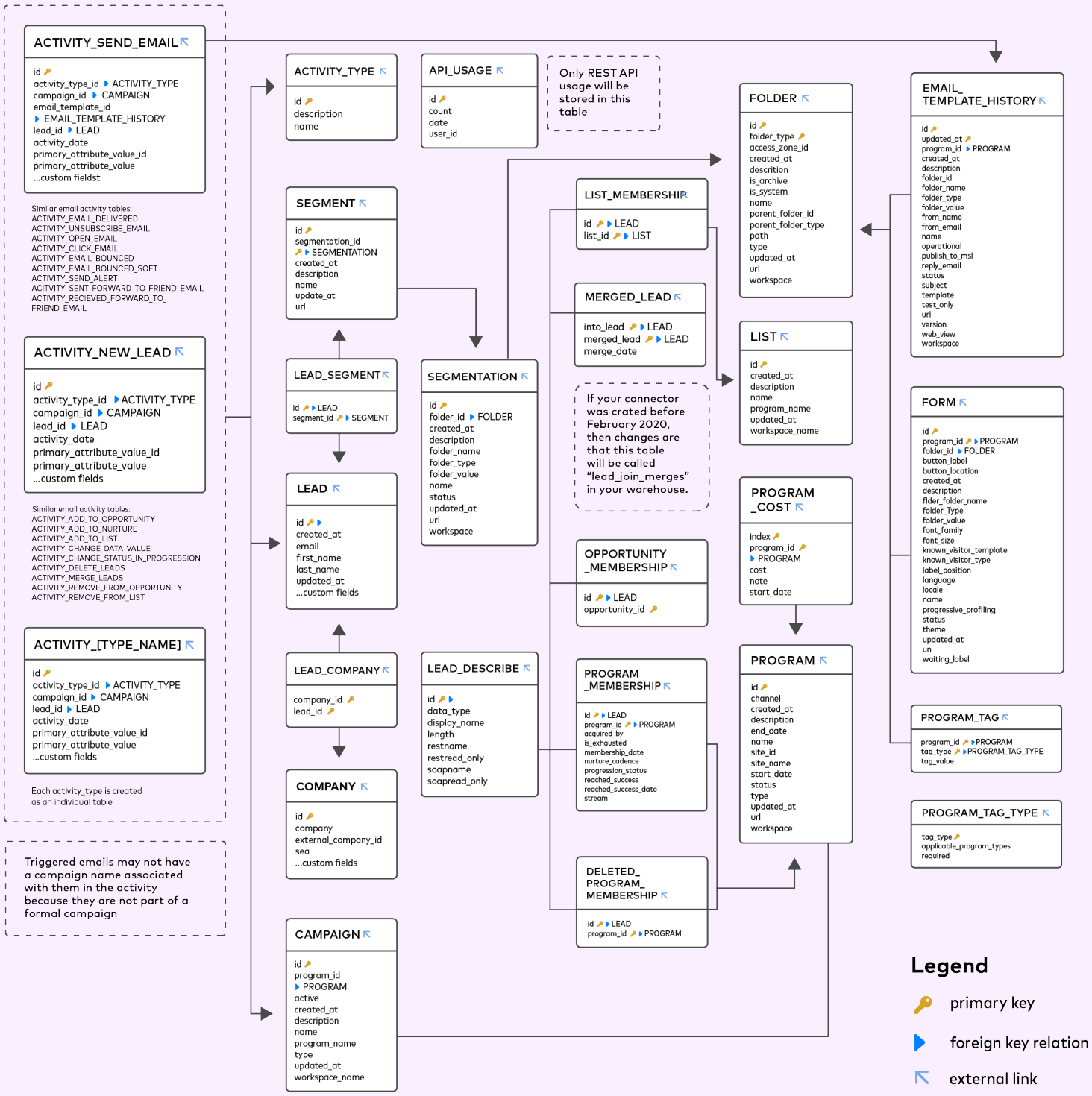
This means that data teams can automate extract-load into a normalized schema. Your never-ending list of data integration projects and maintenance can be replaced by software, freeing up a significant amount of data engineering time for higher-value activities.



Fivetran-enabled workflow with automated access to data and no waiting for data engineering time

[2] We make exceptions where a) normalized data is not available through any API (i.e., ad reporting APIs), b) a semi-structured object is dynamically heterogeneous or c) a deeply structured object is very tangential to core business processes.
[3] For databases, we make a faithful replica of the source schema; for file systems and events, we flatten the first level of fields into columns in a table.

Here's a representative self-describing Fivetran schema for Marketo:



Our normalized schemas also contribute to the success of downstream teams. Those teams can be confident that the data they're accessing is accurate, reducing back and forth between multiple teams and redundant pipeline management.

IMPACT OF NORMALIZATION ON WAREHOUSE COSTS

As you think about the normalization behaviors of your data movement tool or platform, the best-case scenario looks like this:

- Your tool or platform automatically provides thoughtful, well-designed schemas, freeing up engineering time and accelerating analytics
- The tool or platform normalizes your data within its own systems, so the process doesn't drive up your data warehouse costs

If you're using warehouse-native data integration tools like AWS Glue or Azure Data Factory, you'll have to normalize the landed data yourself, a process that can eat up substantial compute bandwidth and increase costs. For example, those tools load all the data, including duplicate records, and you'll need to use transformation compute power to identify and omit the duplicates.

If you're using a third-party data movement tool that offers normalization, it's important to know whether the provider performs the normalization within its own systems or within your data warehouse or destination. If it's the latter, it can get expensive.

PRINCESS POLLY

Anand Bhatt, Head of Business Analytics at fashion startup **Princess Polly**, has found Fivetran schemas for critical ecommerce data sources like **Shopify** and **Klaviyo** to be extremely helpful.

> "Fivetran was a no-brainer. The schema data pool — the table structure that gets pulled from Shopify, and from other data sources like Klaviyo — was definitely easier. The process was also smoother than with other tools.
>
> We acheived our goal of creating value from automation in the analysis — providing metrics, rather than spending time cleaning and moving data."

Anand Bhatt, Head of Business Analytics at fashion startup Princess Polly

At Fivetran, we normalize your data within our own VPC, so you'll never have to worry about data ingestion processes devouring your warehousing compute bill. We've found that many data integration customers struggle with high compute costs from data normalization performed in third-party data warehouses.

Here's what it looked like when a current customer simultaneously used Fivetran and two other data integration providers to replicate the same data into the same warehouse; the other providers used the customer's warehouse instance to normalize the data.
(Figures are for monthly usage.)

Data movement provider	Warehouse credits used	Percentage of compute costs
Provider 1	143.69	61%
Provider 2	69.33	30%
Fivetran	6.74	3%

Look out for this hidden inefficiency when you're evaluating data movement providers.

Idempotence

"Idempotence" is admittedly a **pretty esoteric term**, but in the context of data movement it refers to the ability of a data pipeline to prevent the creation of duplicate data when syncs fail.

For the sake of efficiency, data warehouses load records in batches, so sync progress is recorded by the batch, not the individual record. When a sync is interrupted, it is often impossible to pinpoint the precise record that was being processed at the time of failure. This means that when the sync resumes, the pipeline must start at the beginning of the most recent batch, and some data is processed again.

We can represent a non-idempotent sync failure and "recovery" visually.

First, a four-batch sync is interrupted during the second batch.

Then, when the sync resumes, the pipeline refers to the cursor and restarts the second batch, reintroducing previously synced records.

With an idempotent process, by contrast, every unique record would be properly identified and no duplication would occur.

Sync failures are inevitable, and they can occur at the source, pipeline or destination stages of the ELT process. In every case, however, recovery from failure requires rolling back to some previous cursor and reintroducing some number of records to a destination. Without idempotence, this reintroduction will cause records to be duplicated, undermining the integrity of the data repository.

Data teams have to fix these data integrity issues manually — which means spending hours deleting duplicate data or backfilling missing data, and then reviewing the datafor completeness.

Fivetran engineers idempotence into every data pipeline, so failures do not produce duplicate data or associated downstream problems. Fivetran pipelines are essentially self-correcting, eliminating the need for data teams to engage in time-consuming data integrity forensics.

Incremental syncs

Another key feature of an efficient data pipeline is the ability to update incrementally. The alternative is to sync entire data sets every time you update from a source. While full syncs are necessary to capture all your data initially, they're inappropriate for routine updates because they:

1. Often take a long time, sometimes to the tune of hours, days or even weeks for exceptionally large data sources. Full syncs thus cannot provide timely or real-time updates, and they may time out, impacting data integrity and access.
2. Can bog down both the data source and the destination, using resources otherwise needed for operations and analytics, respectively.
3. Consume excessive network bandwidth.

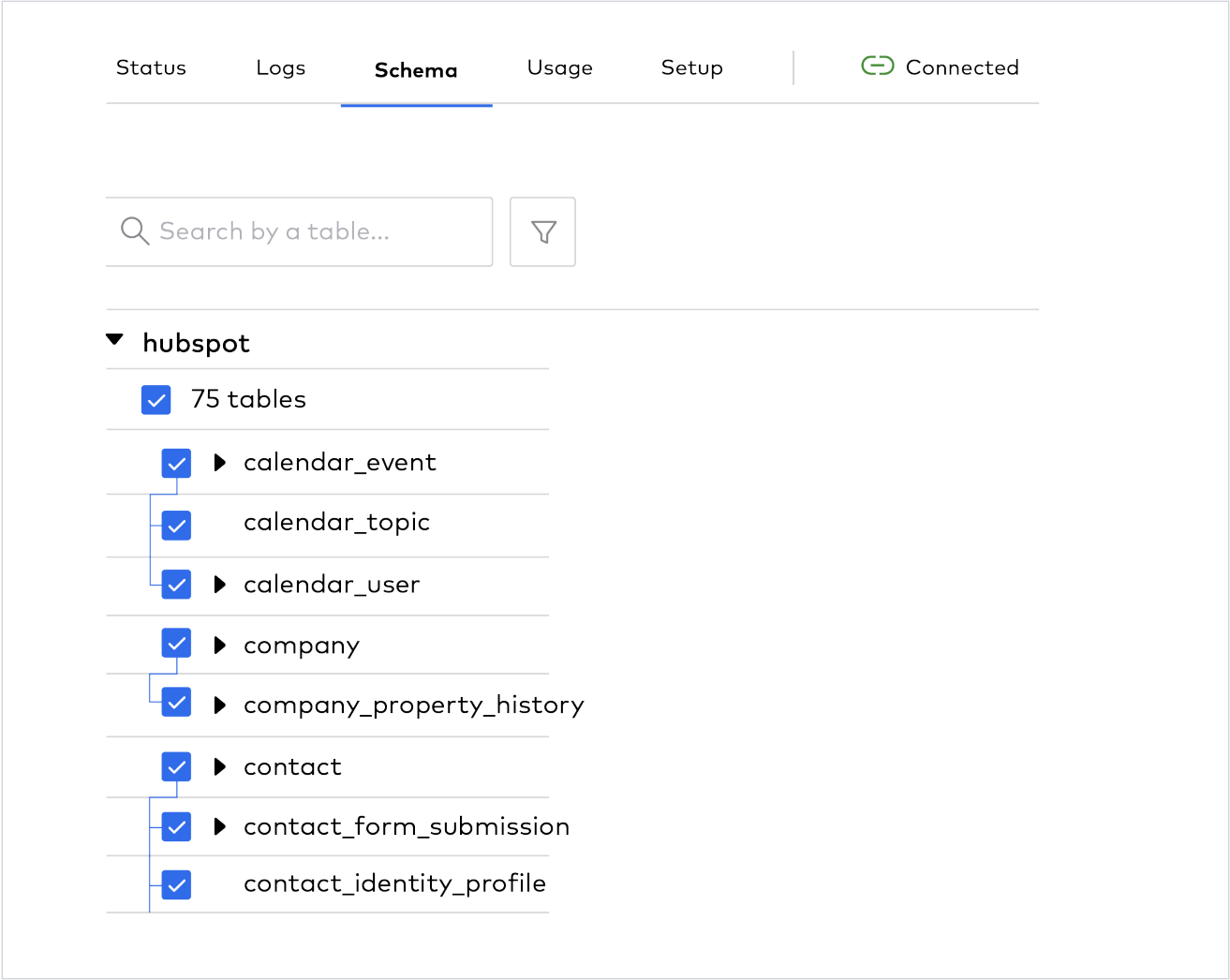
Keep in mind that the difference in data volume between a full and incremental sync can be one or two orders of magnitude, with resulting increases in data warehouse costs. And as data sources grow and add records, the problems identified above grow as well.

Performing incremental syncs requires identifying changes made to a previous state of the source, a practice often referred to as **change data capture**. Make sure your data movement provider syncs data incrementally — and does so in a way that **solves the common problems associated with incremental syncs**. Otherwise, the inefficiencies caused by those problems will be passed along to your data team.

Data selection

Your team might not be interested in analyzing the data in every single table an ELT provider syncs by default into your warehouse or destination. If those analytically irrelevant tables are large, they have the potential to significantly increase your data integration and warehousing costs. We commonly see customers looking to avoid syncing logging tables, notification tables and external data feeds that have little or no historical or analytical value to the end user.

Efficient ELT platforms will allow you to granularly de-select tables and columns that are unnecessary for analysis. Fivetran makes it easy for your team to do this.



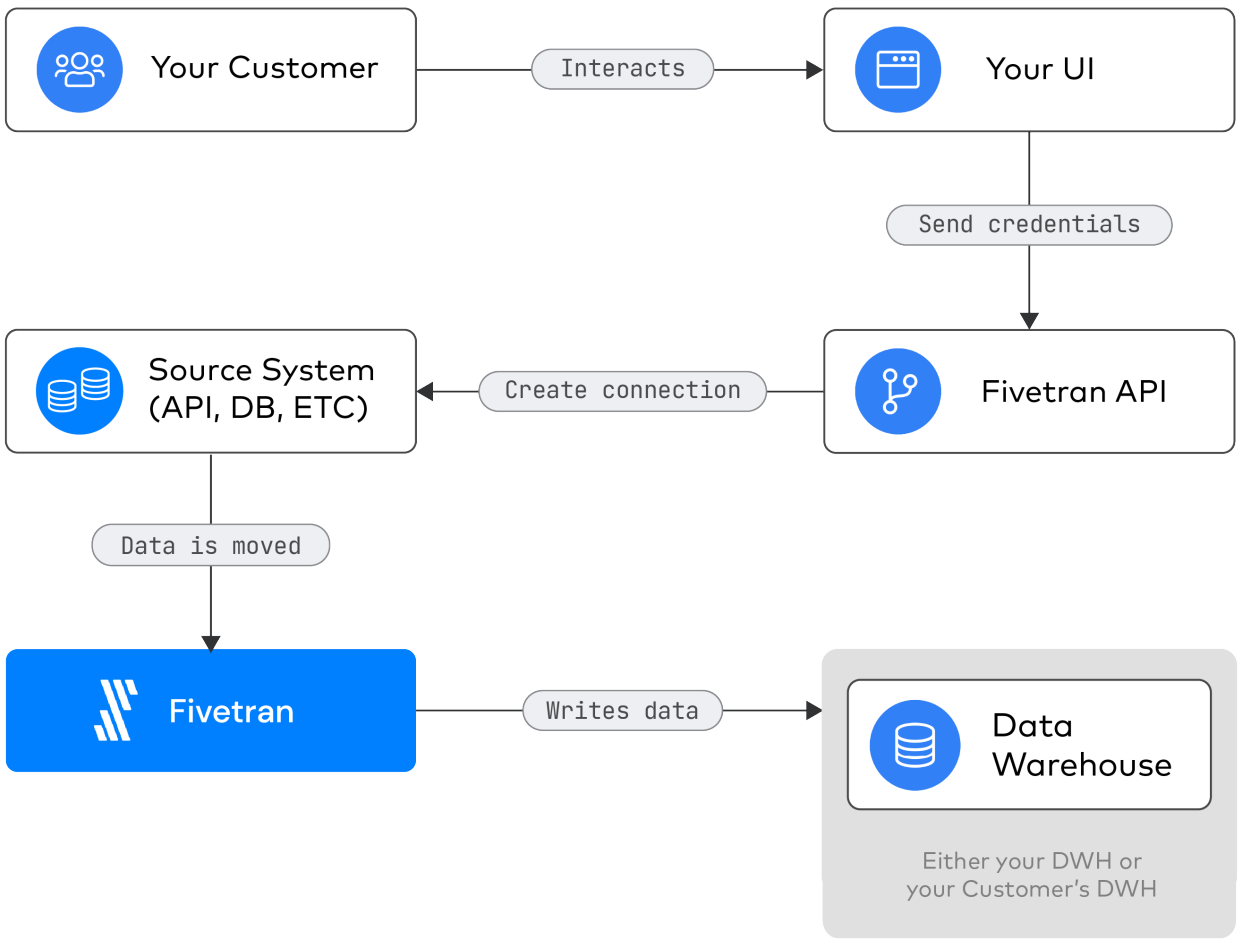
Programmatic pipeline management

The ability to manage data pipelines via an ELT provider's API, not just its user interface (UI), can hugely reduce your team's workload — and it's an essential component for anyone trying to build data solutions at scale. If you need to perform the same action multiple times, an API helps you reduce errors and inefficiencies, saving you time and reducing stress. **Look for ELT providers that support programmatic pipeline management via API.**

The Fivetran API offers an efficient, consistent way to perform bulk actions, communicate with Fivetran from a different application, and automate human-led processes using codified logic. Engineers can create and edit any number of data connectors at once, while bulk actions can be made to control user access and permissions and even provision any quantity of new groups of users.

Data teams that can programmatically design workflows and eliminate low-value repetitive actions will have far more time for business-critical tasks and advanced data analysis. For example, **we've helped customers use the Fivetran API to create connectors to hundreds of databases — and in one instance over 1,000 — saving them many hours of work in the process.**

Here's an illustration of just one of the API use cases we support — **data consultancy Untitled Firm using the Fivetran API** to connect to its customers' data and create powerful analytics applications:



V.

How an ELT platform can accelerate analytics

Ideally, an ELT tool or platform takes an integrated approach to the "T," making data transformation as simple, powerful and seamless as possible within its own UI. Its transformation offering should follow evolving best practices in the analytics world, providing CI/CD features, prebuilt data models and data lineage graphs.

The approach should also be accessible. SQL has become the de facto coding language of data transformation, so a good ELT tool should be SQL-based, ensuring all analysts across the organization can leverage it.

Collectively, these best-practice capabilities allow your team to create scalable and powerful transformations much faster. When your analysts have the tools they need to do their jobs more easily — using the skills they already have — they're more efficient.

If your team spends less time doing things like poring over API documentation, conducting root-cause analysis and manually coordinating data syncs and transformation runs, they'll be much more effective supporting business decision-making.

Here are the capabilities you should look for in terms of integrated transformations, many provided by dbt™, the leading SQL-based transformation tool. Fivetran's transformation layer integrates with dbt and provides additional functionalities that accelerate your analytics.

Fast, collaborative and transparent modeling

VERSION CONTROL

dbt Core offers a full change history of your transformation models, making troubleshooting and validating data integrity far more efficient. You can easily revert transformation models back to previous versions if needed and compare changes over specific periods of time.

Branching and merging capabilities facilitate collaboration, allowing team members to work concurrently while keeping multiple work streams independent. You can check for conflicting changes while individual branches are being merged, and easily request a review during this process.

This also ensures that the main branch where transformations live always supports production-quality code — transformation models won't "go down" while they are being worked on or new models are being tested.

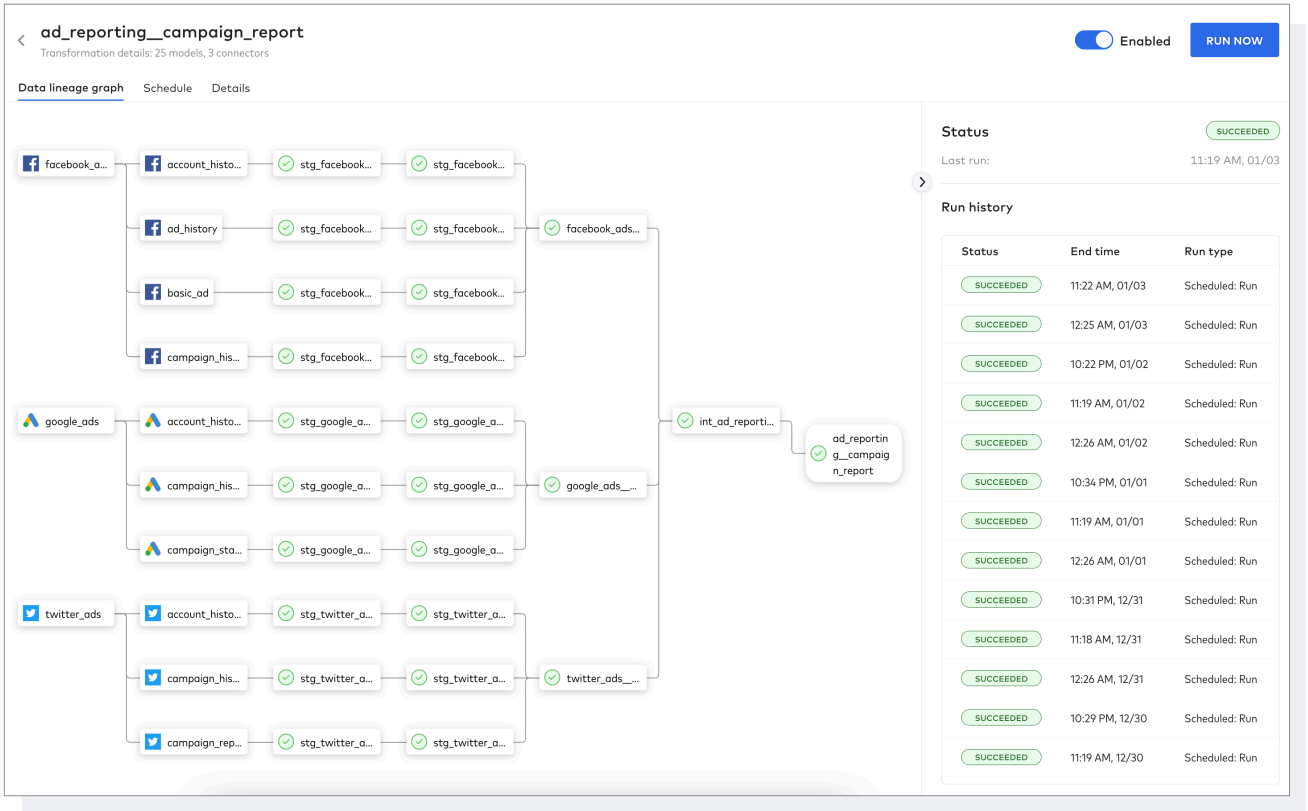
CI/CD is rapidly becoming a best practice; along with dbt, Fivetran is committed to following it so that our customers can scale their transformations effectively.

PREBUILT DATA MODELS

Fivetran has created an extensive **library of open-source data models** that can be readily applied to supported data sources. Built on top of common table schemas from our connectors, Fivetran data models automate much of the dbt development required to make your data analytics-ready.

Using Fivetran prebuilt data models reduces the need to manually code SQL, saving your data team weeks of time. You'll be able to focus on deeper analysis and generate insights faster instead of doing foundational data modeling.

Our **ad reporting package**, for example, rolls up ad spend, clicks and impressions across Apple Search Ads, Facebook Ads, Google Ads, LinkedIn Ad Analytics, Microsoft Advertising, Pinterest Ads, Snapchat Ads, TikTok Ads and Twitter Ads, offering a unified view of your ad efforts:



Developed with data team efficiency in mind, Fivetran data models:

- **Automate the upstream transformations** that lead to the output model you want to generate, so you only have to focus on the end use case for the data — we manage all of the staging and intermediary steps
- **Have rich, built-in metric definitions** based on an integration between Fivetran and dbt Semantic Layer, ensuring all downstream users understand the data they're analyzing
- **Detect upstream connectors automatically,** meaning your transformations will be in sync with your extract and load
- **Are extensible,** so your team can include logic that meets your specific business needs

mighty.digital

Analytics and strategy consulting firm Mighty Digital has helped several companies find efficiencies and optimize their data-driven performance using Fivetran data models.

> "The data models provide two primary benefits: parsed, defined and described data models as well as ease of use.
>
> The first saves a ton of time while working with the source data.
>
> Second, the ease with which you set up the source data integrations, connect pre-configured data models for that source and then integrate it all into a set of transformations is just outstanding.
>
> Overall, the process of implementing Fivetran and dbt together was very straightforward — sometimes I wondered if these two technologies were created for one another."

Vladyslav Hrytsenko, CTO and Senior Technology & Strategy Consultant

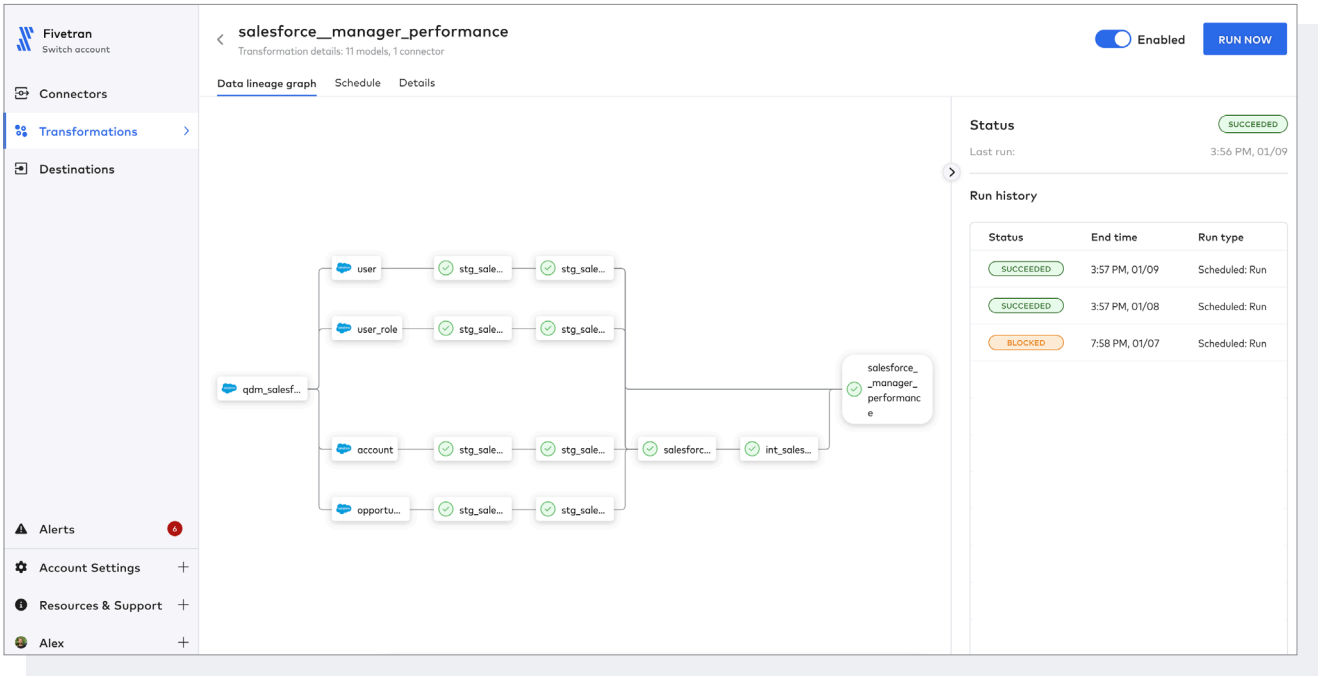
MODULARITY

With dbt, you can reproduce a transformation across multiple models and handle the dependencies between models. For example, if you have a "product" model that joins a product table to a category table, you can reuse this model across all relevant data sources — minimizing your need to build code from scratch.

DATA LINEAGE GRAPHS

Our transformation offering, **Fivetran Transformations for dbt Core™***, automatically generates data lineage graphs, which provide a full visual representation of a transformation from source to output model, including status and logs. Visually exposing complete workflows makes it easy to track the flow of data, monitor performance and quickly debug models in the pipeline to avoid data latency.

Here's an example of a data lineage graph in the Fivetran UI:



Managing and scheduling transformations

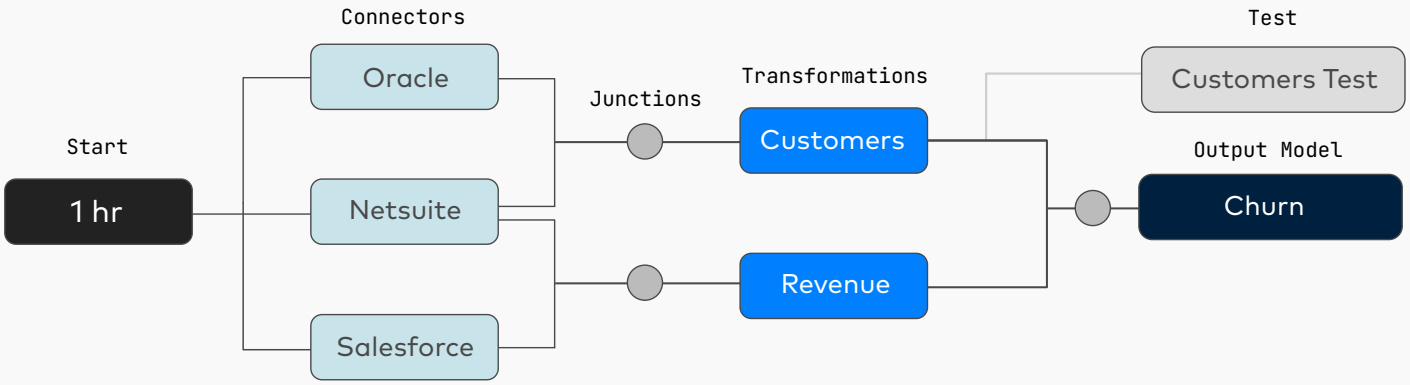
INTEGRATE AN EXISTING DBT PROJECT

Getting started with Fivetran Transformations for dbt Core* is **as easy as connecting a Git repository** with an existing dbt Core project in it. After Fivetran successfully connects with the repository, you have the option of choosing which models you want us to manage and how you want to schedule them. Fivetran will constantly sync with the Git provider, ensuring that all updates are reflected in your models. This reduces your maintenance time by ensuring your data is always modeled most accurately.

If you're new to dbt, our **Professional Services** team can quickly get you up to speed. From installation and implementation of key components, such as your git repository and Fivetran integration, to transformation development, we'll make sure your adoption of dbt Core is fast and successful, with minimal impact on the bandwidth of your team.

SCHEDULE AND MANAGE TRANSFORMATIONS IN THE FIVETRAN UI

After we connect to your Git provider, we run your dbt models in your destination according to a schedule that you choose in the Fivetran dashboard. We sync your dbt models from your Git provider on a set schedule, or to match the frequency of your data pipelines to ensure we're up to date.



Importantly for teams looking to save time and keep data fresh, Fivetran offers what is known as "fully integrated" scheduling — by default, your new transformations are synchronized with their associated connectors. Because Fivetran manages the extract and load, we know exactly when to orchestrate your data model transformation, giving you full automation while reducing latency. Your team won't need to manage a third-party tool or manual processes.

QUICKSTART DATA MODELS

To further streamline your ELT process, Fivetran offers Quickstart Data Models, which allow you to use our prebuilt data models without building your own dbt project — meaning you manage the entire ELT pipeline without leaving the Fivetran platform or writing any code. Just set up your Quickstart Data Model and we'll create the dbt project and transformation for you, while also providing our other transformation features. You'll be able to view, manage and edit your entire pipeline right from your UI.

SpotOn

Leading software and payments company **SpotOn** was able to find scale turning their clients' transactional data — hosted in 30 MySQL databases — into client-facing reporting with Fivetran Transformations for dbt Core.

Prior to Fivetran, SpotOn transformed their data using manual stored procedures that grew to over 2,000 lines of code. By implementing dbt Core, they were able to modularize their modeling and orchestrate the process automatically.

> "By integrating Fivetran and dbt, Fivetran becomes not only the ELT tool but also the orchestration tool, which is great for reliability, scalability and dev time.
>
> Fivetran gives us a centralized, one-stop shop for our entire ELT pipeline — no third-party tools required. That made our processes more efficient and cut down on monitoring."

Tom Gilbertson, Project Manager for Data and Analytics at SpotOn

CONCLUSION

VI.

Test efficiency benefits before you buy

Most modern ELT tools and platforms offer free product trials, which can give you and your team a firsthand sense of how they'll contribute to your efficiency and performance across the functional areas we've discussed here. We highly recommend testing out tools and platforms before you commit to making them part of your stack — you'll likely discover hidden inefficiencies you might not have considered.

If you **sign up for a Fivetran trial,**
you'll have 14 days of free access to data connectors for 300+ sources.

The trial doesn't begin until your initial historical sync has completed, and from there you can explore the capabilities we've covered, including deeply researched ERDs, automated transformationsand prebuilt data models.

Fivetran has multiple pricing tiers — including a free plan for smaller or more cost-conscious organizations — so you'll be able to limit your initial financial commitment if necessary. You'll also benefit from free historical syncs and priority-first syncs, which allow you to access your most recent data without having to wait for the initial sync to complete.

Fivetran

Fivetran automates data movement out of, into and across cloud data platforms. We automate the most time-consuming parts of the ELT process from extracts to schema drift handling to transformations, so data engineers can focus on higher-impact projects with total pipeline peace of mind. With 99.9% uptime and self-healing pipelines, Fivetran enables hundreds of leading brands across the globe, including Autodesk, Conagra Brands, JetBlue, Lionsgate, Morgan Stanley, and Ziff Davis, to accelerate data-driven decisions and drive business growth. Fivetran is headquartered in Oakland, California, with offices around the world.

For more info, visit Fivetran.com.

Start your free trial

©2023 Fivetran Inc.