

Modernize your data transformation skills and accelerate your career

Discover how modern data transformation tools are speeding data analysis efficiency — and why you need them to jumpstart your career.

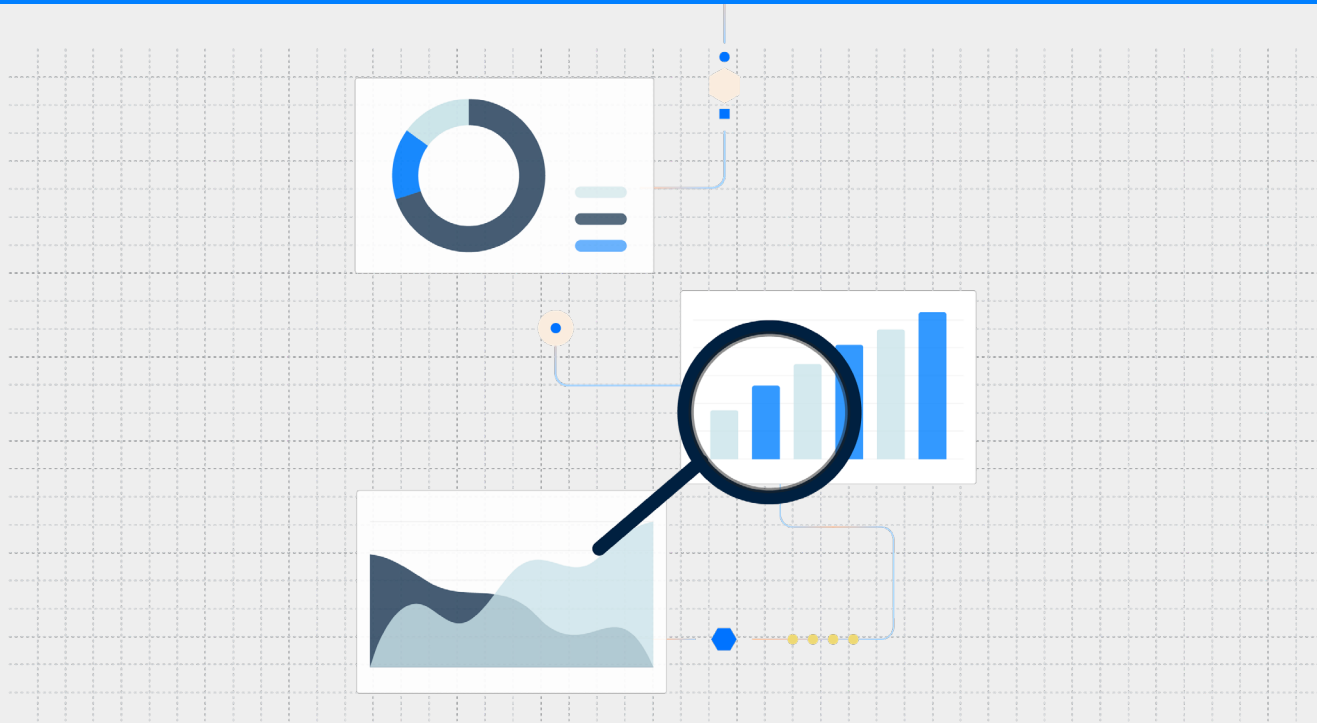


Table of contents

Part 1: The state of data and analytics	3
Part 2: Companies are transitioning from ETL to ELT	4
How ELT beats ETL	4
A few considerations about GUIs	5
ELT is the future of data	6
Part 3: Why SQL will remain an in-demand skill in 2023 and beyond	6
How SQL can modernize	7
Part 4: dbt: A new way to transform data and build pipelines	7
The advantages of using dbt	8
What this means for you	9
Choosing the right dbt product for your data environment	9
dbt Cloud for collaborative transformations	9
Part 5: The ultimate data transformation solution: dbt + Fivetran	10
How Fivetran supports data transformation	10
Using Fivetran Transformations for dbt Core	10
Autodesk: Transformations happen within minutes with Fivetran + dbt	11
Fivetran + dbt speeds up data transformation process for Phenix	12
Advance your career in data analytics with modern tools and processes	12

Part 1: The state of data and analytics

Businesses are collecting data from every corner of the organization to help them make intelligent and informed business decisions. Yet, most of this data is being underutilized.

Most companies simply don't have the time or capabilities to analyze all of the data they collect. [In a recent survey](#), 68 percent of data professionals shared that they could extract more insights from existing data if they only had the time, and 44 percent said that even crucial data is not yet usable for decision-making. As a result, companies are missing out on key opportunities to drive their business forward and optimize their KPIs — using data that could help them do more with less, when they have to pull back on budgets.

However, a new approach to unlocking data has emerged. The modern data stack (MDS) offers greater speed, scalability and flexibility in ingesting and transforming data for analysis. Built with analysts and business users in mind, the MDS is a suite of cloud-based tools, including:

- Data ingestion/integration tool
- Cloud-based warehouse or data lake
- Data transformation tool

[Gartner](#) predicts that by 2024, 50 percent of data and analytic deployments in the cloud will be based on a cohesive cloud data ecosystem rather than manually integrated point solutions.

A modern data stack offers an automated way of keeping up with the business and supporting decision-makers with mission-critical and consistently up-to-date information. It also allows data analysts or solution engineers to improve their skill sets with emerging technologies.

As companies strive to get the most from their data, the intense demand for data analysts will only continue to grow. [The U.S. Bureau of Labor Statistics](#) estimates jobs for data analysts will increase 23 percent between 2021 and 2031 – which is much faster than the average for all occupations.

Despite the growing demand for data analysts, global and economic circumstances mean competition is stiff. Employers are looking to data teams to accelerate the transformation of data more effectively and more efficiently.

This means you, as a data analyst, have the opportunity to not only directly impact your company's goals but also to propel your career. This ebook will show you how to modernize the data transformation process using the latest tools and technologies to increase impact and forge a successful career path.

Part 2: Companies are transitioning from ETL to ELT

In one recent survey, respondents reported spending an average of **45 percent of their time on getting data ready** (loaded and cleaned) before they could use it to develop models and visualizations. The shift from ETL (Extract-Transform-Load) to ELT (Extract-Load-Transform) is driven by a desire of most organizations to be more efficient with their data processes. Whittling down the time it takes to transform data has become one of the biggest priorities for most organizations.

How ELT beats ETL

ELT decouples extraction and transformation, moving data into a data warehouse before applying business logic to transform it into data models for analytics. The benefits of this approach versus ETL include:

- **Shorter turnaround timelines**
Because analysts can perform transformations within the data warehouse environment without needing to rely on data engineers, it shortens the turnaround times for all analytics projects and allows for speedier delivery of insights.
- **More affordable, easier to use and more practical**
The explosive growth of cloud-based services no longer requires businesses to manually build and maintain data pipelines. Companies can afford to store untransformed data in cloud-based data warehouses.

- **Increased data flexibility**
Moving transformations to the warehouse means you don't have to rebuild pipelines when query parameters change. This makes your insights more flexible and saves you re-engineering time/cost.

Importantly, ELT eliminates the two points of failure within the ETL process — changing upstream schemas and downstream data models from impacting extracting and loading. This makes ELT a simpler and more robust approach to data integration.

Automate your data pipelines with Fivetran and dbt™

Unlike the multitude of steps, tools and technologies characteristic of ETL, the flow for Fivetran is simple, effortless and can be entirely done by an analyst:

1. Set up a data warehouse within Fivetran
2. Hook up Fivetran connectors and begin syncing
3. Write and schedule dbt transformations to create the data model you need
4. Build dashboards and reports

This approach is accessible, built on simple authentications and SQL skills.

A few considerations about GUIs

When it comes to data transformation, there are many vendors on the market to choose from. However, many of these vendors rely on GUIs (graphical user interface) and non-cloud-based processes to support ETL.

The appeal of GUIs is that they offer users pre-designed components to drag and drop onto a canvas to visually build out data flows. While GUIs look and sound easy to use, they can often be rigid and unintuitive.

Here are some of the downsides to GUIs:

- **Vendor lock-in**

GUIs are proprietary to each vendor, which means that jobs are not transferable from one application to another. So, if you or your company wants to switch the tool or method you're using, you'll need to rebuild those jobs from scratch.

- **Limited capabilities**

Because there may be actions you want to take that aren't natively supported, GUIs can limit what you can do.

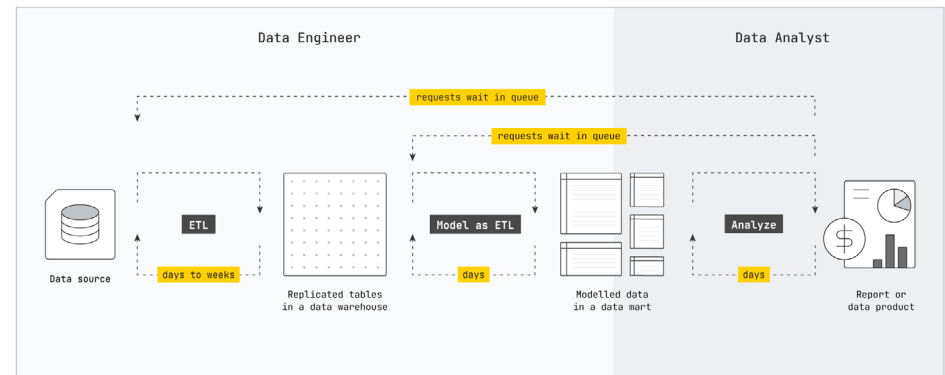
- **Non-transferable skills**

A GUI requires learning specific skills to use it and those skills are only applicable for that GUI. So the skills aren't as transferrable in the job market as they would be working in a typical straight-up code environment.

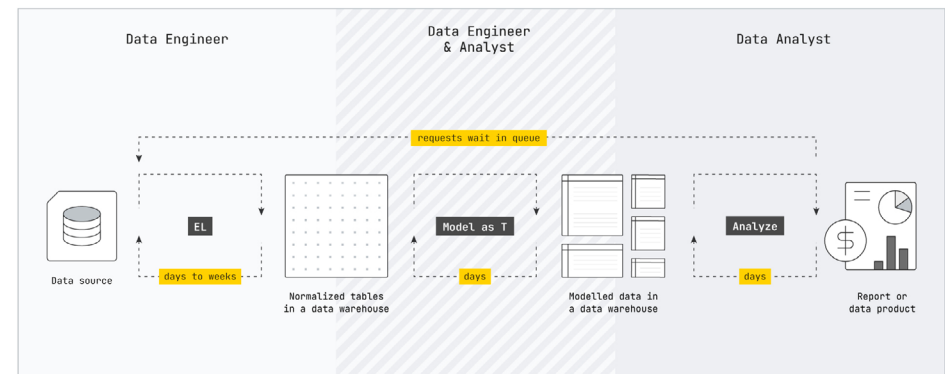
- **Human capital investment**

Learning a new GUI can take weeks to months. You often have to hire engineers who are specifically skilled in that platform. And, for those who want to learn, the learning curve is steep. These factors delay the delivery of data analytics initiatives and limit who you can collaborate with.

Legacy workflow



Modern workflow



ELT is the future of data

ELT simplifies data integration, results in lower failure rates, allows for flexible scaling and moves the transformation process to the warehouse, where you can apply such skills as SQL to achieve data transformation (we'll get into why this is important in the next chapter). All of this is to say that ELT is a simpler, faster and more affordable data pipeline process — and the better option for the vast majority of organizations looking to get more valuable and actionable insights from their data faster.

→ See how Fivetran can help you with modern ELT

Part 3: Why SQL will remain an in-demand skill in 2023 and beyond

When it comes to skills that every data analyst or solutions engineer should have — SQL typically sits at the top of the list. Given that almost every business uses systems with relational databases and SQL is the programming language required to manage and interact with these databases, you should have a solid grasp of SQL.

SQL continues to remain the standard for working with data within relational databases for several reasons:

- **Quickly handles large volumes of data**
SQL is typically quite fast. One SQL query can fetch results from millions of rows within a minute, allowing you to work with large volumes of data.
- **Repeatable and scalable**
SQL queries you write are repeatable and scalable, meaning you can write a query once and reuse it again and again. Even if you add new tables, you can use the previous query by making a few minor changes. And because most databases are now cloud-based, they can support scalability for quickly adding new data.
- **Greater flexibility**
Unlike the prescribed transformations for components, as long as it is possible in SQL, you can transform the data to solve for unique business logic.

How SQL can modernize

SQL is a fantastic tool — understood by nearly anyone with "data" in their title. But without a way to make it reusable and shareable, it can actually slow teams down over time. Some areas in which SQL can modernize include:

- **Transparency and visibility**

Teams often write transformation scripts in silos — moving too quickly to share, test, or de-duplicate. However, locally developed code creates major risks for the team and organization at large, slowing time to development, compounding data quality issues, and making audit and compliance nearly impossible.

- **Built-in automation**

Transformations in SQL are often coded from scratch for each use case. New data sources and tables often require manual work to model - even if you're building from a similar baseline. This means that resource time is being used inefficiently and there are opportunities for error building with each new line of code.

- **Lack of built in documentation, version control, and CI/CD**

Without best practices like documentation, version control and continuous integration/continuous delivery (CI/CD) built in, these all have to be done manually and can be time-consuming and error-prone. It also means that you'll likely need additional tools in your data stack to meet your goals (including orchestration, scheduling, data lineage, etc.).

While SQL is still the de facto language for data analysts and data modeling, that doesn't mean there isn't room for improvements. Luckily, new tooling has helped to alleviate these concerns and address the data analytics challenges that data teams face. In the next chapter, we'll discuss these new tools and how they've evolved the data transformation process.

Part 4: dbt™: A new way to transform data and build pipelines



There are many tools on the market to help with data transformation, but dbt by dbt Labs™, which is an open-source tool that simplifies data transformation, stands out for its ability to expedite the process of transforming data and building data pipelines.

dbt is built on the SQL skills you have but modernizes your deployment. Historically, only data engineers had the technical capabilities to perform these activities. But, by using dbt, you can take on some of the engineering work, heightening your value within your current organization and the marketplace.

dbt evolved from the belief that data analysts should adopt similar practices and tools to software developers. It was also founded on the belief that code, not GUIs, is the best abstraction to express complex analytic logic.

The advantages of using dbt

There are several distinct advantages of using dbt as your data transformation tool:

- **SQL-based**
Because SQL has long been the language of choice for analysts developing transformations and performing analysis, you don't need to learn a new language or tool.
- **SELECT statements**
dbt allows you to use SQL SELECT statements. This is an important feature because even within SQL, managing transformations usually requires in-depth knowledge of more complex queries. By allowing you to express analytical logic more simply, dbt makes it easy to turn data into business intelligence.
- **Automated dependency handling**
With the reference function in dbt, once analysts define a relationship between a set of data assets (i.e., tables or views), dbt automatically creates them in the right order to enforce that relationship, every single time. Combined with automated documentation, this improves the visibility of up and downstream dependencies without manual logging, and allows teams to work faster.
- **Modular Development**
dbt makes SQL modular - rather than cutting and pasting frequently used snippets of SQL, dbt allows the use of SQL macros and dbt packages, which can automate many common tasks. This allows analysts to avoid repeating frequently-used

statements, saving time, preventing errors, and making code more legible.

- **Open-source**
dbt is open-source, which means it is often more flexible and adaptable than proprietary software. Development is faster and more innovative within an open-source environment, in part because you can collaborate with a community of experts.

Another area that dbt shines in the data transformation process is its capability to provide software engineering best practices such as:

- **Automated documentation:** dbt will generate documentation and diagrams of the SQL models you have created and dependencies between the models.
- **Automated testing:** You can define and schedule tests to be performed after each transformation, whether it's checking that a column contains no nulls or ensuring referential integrity between tables.
- **Collaboration:** Large teams can contribute to the transformation process because all transformations are expressed as code and can thus be committed to a version control system like Git.

These best practices allow you to spend less time testing so you can speed your time to development and ensure you deliver trusted data across the organization.

→ Ship better data faster. Create a free dbt account

What this means for you

With this new skillset, you can amplify your impact now and set yourself up for success in future endeavors.

- **Current impact:** With dbt, you as a data analyst can expedite the time to insight. Using the SQL-skills you already possess, you can more effectively build data transformation pipelines at scale. Increased collaboration and built-in version control ensures SLAs are met and data latency is reduced. This makes you a more trusted insights-generating partner to all stakeholders, who get faster and more reliable insights, increasing their ability to optimize your desired business outcomes with the power of data.
- **Future impact:** Being able to demonstrate that your actions not only made your company more efficient, but also more impactful, makes you more attractive in a competitive job market. And, you bring with you a skill that you can apply immediately to any data stack, ensuring you have a big impact on any new team you can join.

Choosing the right dbt product for your data environment

dbt has two main products. dbt Core™ is an open source tool that includes everything technical teams need to transform data.

It provides you with the ability to roll your own ELT pipelines, SQL compilation logic, Jinja templating, database adapters, testing frameworks and documentation.

dbt Cloud™ is great for teams that want to move quickly in a highly coordinated way. It is fast and reliable, has a user-friendly IDE, allows for job scheduling, centralized logging and alerting, and provides CI/CD with Slim CI, among other things.

When assessing which dbt product is right for your organization, consider your team structure and skills. No matter your choice, dbt will help you modernize your approach to data transformations.

dbt Cloud for collaborative transformation

dbt Cloud provides the most fully-featured experience for teams using Fivetran and dbt:

- Enable analysts to safely contribute to the transformation workflow with an intuitive IDE for version control and orchestration
- Run only the jobs you need using "Slim CI" with the dbt Cloud scheduler
- Optimize model builds with a visual monitor to quickly identify long-running models
- Leverage parallel run slots to speed up time to delivery
- Power consolidated lineage and metrics with the Metadata API and dbt Semantic Layer

→ [Learn more about the dbt Cloud](#)

Part 5: The ultimate data transformation solution: dbt + Fivetran

How Fivetran supports data transformation

While dbt helps you modernize the approach to data transformation, it's only part of the process. Before you can start seeing value in a tool like dbt, you must first ensure that the rest of your modern data stack is in place. This includes a fully-managed data pipeline provider to ingest a copy of your source data into your cloud-based destination. This is where Fivetran comes into the picture.

"dbt focuses on doing one thing and doing it incredibly well — handling the transform step in an ELT pipeline. Fivetran can efficiently handle the extract and load stages, and as neither tool requires additional infrastructure, they can be combined to create a streamlined but flexible pipeline."

Vlad Hrytsenko, CTO at Mighty Digital |  mighty.digital

Fivetran offers 200+ prebuilt data connectors to help make it easy to automate ingesting and centralizing all data sources, including applications, databases, events, files and functions. It automatically creates and manages schemas and data types appropriately for your destination. Tables are immediately ready to query in your data warehouse or data lake.

Fivetran has authored 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, these dbt packages automate much of the development to make your data analytics ready - giving you a head start to turning data into insights.

Using Fivetran Transformations for dbt Core*

Built as a free, native integration, Fivetran Transformations for dbt Core allows you to run your dbt Core transformation models directly from our dashboard. Key features of this integration that make your data transformation processes faster and more efficient include:

- **Integrated scheduling** that automatically triggers model runs following the completion of connector sync loads in your warehouse
- **Data lineage graphs** that let you monitor data movement from connector load through transformation steps to final output model
- **Alerts and notifications** that help simplify troubleshooting

*dbt Core is a trademark of dbt Labs, Inc. All rights therein are reserved to dbt Labs, Inc. Fivetran Transformations is not a product or service of or endorsed by dbt Labs, Inc.

"Fivetran Transformations makes Fivetran a full EL and T platform, rather than just EL. It ensures your Fivetran data connectors and transformations happen in the right order, at the right time."

David Jayatillake, Head of Data at Metaplane |  metaplane

→ Watch our video series to get started with our integration

Autodesk: Transformations happen within minutes with Fivetran + dbt

The [Autodesk](#) Construction Services data team's main challenge was their data extraction, storage and transformation process. Extraction required daily monitoring, with an analyst going into the system to make sure no data was missing and no schema changes from upstream affected downstream workflows. There were over 20 downstream models that the business had to ensure were sound. Limitations on parameters and backfill functionality meant that there could be days worth of data missing that required manual backfill, and transformations could take up to 23 hours.

By building a modern data stack consisting of Fivetran, Snowflake and dbt, Autodesk could:

- Create highly denormalized tables for source standardization
- Adjust warehouse sizes
- Increase concurrency
- Implement processing of JSON and list transformations in the warehouse

"With Fivetran we can do automatic schema migrations, so there are no stoppages in the flow of data. We have tests to troubleshoot specific areas as they may come along, but it doesn't stop our end users from being able to use our reporting tools. The design is intuitive and simple, the connector coverage meets most of our tooling needs and customer service is great."

Evin Anderson, Data Engineering Manager at Autodesk |



The results:

- Pipeline maintenance has gone from three to five percent of analyst time to less than one percent
- All transformations are done in-warehouse
- The extraction process no longer requires daily monitoring thanks to automated email alerts if a connector is delayed or a sync fails
- Transformations happen within minutes, helping analysts provide insights more quickly

Fivetran + dbt speeds up data transformation process for Phenix

Phenix, a “tech for good” company, launched a web platform where unsold food products could be made available at reduced prices. The challenge was to find an effective way to turn data captured on the website and in a range of applications into actionable insights.

Phenix uses Fivetran and dbt to power these insights. Fivetran consolidates and centralizes all of the data sources in their data warehouse. It then uses dbt for pipeline tests and checking for duplicates, which has sped up the process of bringing clean and accurate data to dashboards in Looker, the company's BI tool.

The result is data modeling that delivers key metrics to help make Phenix more effective, such as re-valuations of unsold products, rates of breakage that leads to waste, and the number of tons of products saved and equivalent impact on CO2 emissions.

“We don't have to spend any time updating the APIs. Whenever the source changes or a new field or column is introduced, it automatically changes in our own database. Fivetran manages it all for us. It runs in the background, which means we can focus on the transformation and modeling of the data — areas where we really add value.”

Andy Barakat, Data Engineer at Phenix | 

Advance your career in data analytics with modern tools and processes

With Fivetran + dbt, you can:

- **Automate**, meaning you spend time on more impactful work and less time on maintenance and build
- **Collaborate with your team**, while inviting new team members to participate in the insight generation process
- **Become a more trusted insight partner for your stakeholders**, rather than just a pipeline maintainer, making you more visible and important
- **Save your company money**, while increasing their efficiency.
People who save money and drive more impact get rewarded

You become more desirable because you have the skills needed to be an impactful part of an efficient team. And you can bring work with you from previous jobs and rebuild using Fivetran's standard data schemas and data models. This means your impact will be felt immediately. With a modern data stack that combines Fivetran + dbt, you can save time, transform more data faster and develop a strong set of marketable skills.

→ Sign up for a 14-day free trial of Fivetran.



Fivetran is the global leader in modern data integration. Our mission is to make access to data as simple and reliable as electricity. Built for the cloud, Fivetran enables data teams to effortlessly centralize and transform data from hundreds of SaaS and on-prem data sources into high-performance cloud destinations. Fast-moving startups to the world's largest companies use Fivetran to accelerate modern analytics and operational efficiency, fueling data-driven business growth. For more info, visit [Fivetran.com](https://fivetran.com).