# Build vs. Buy: An Illustrated Guide

Consider these comparisons before you try to build your own data pipeline

Build vs. Buy

The average business today uses [well over 100 software apps](#), many of which contain valuable insights about an organization's operations. Your company is likely on the way to using just as many apps, if not more, and you'll need a solution to integrate all of the data your apps produce. As you pursue data integration, be sure to consider the benefits an automated, off-the-shelf solution may bring you. Weigh the following comparisons:

Time and Effort
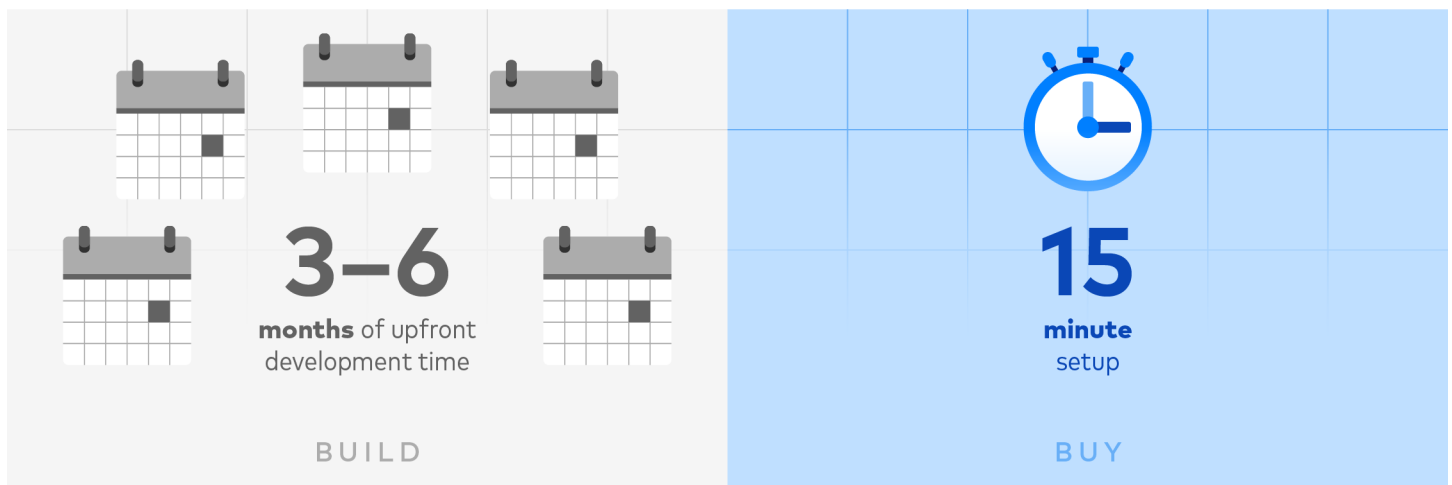
Building your own pipeline is a significant time commitment. Based on our customers' experiences, it can take between three to six months to set up a basic pipeline.

By contrast, an off-the-shelf solution with prebuilt connectors can be set up in a matter of minutes.

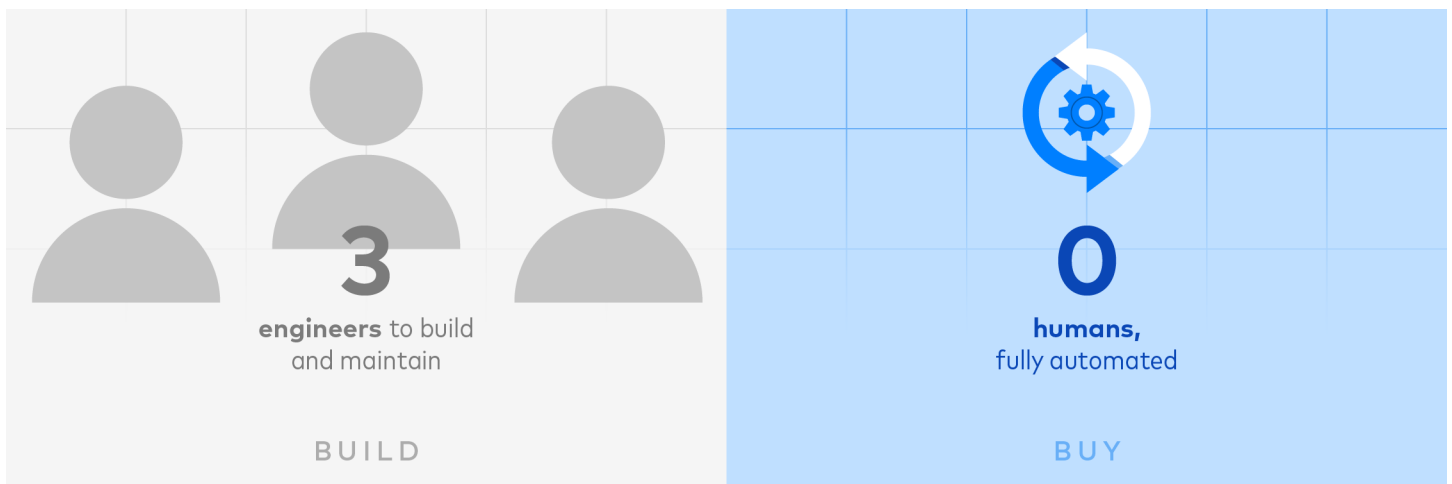| Laborious, multi-step production process	Easy setup
BUILD	**BUY**
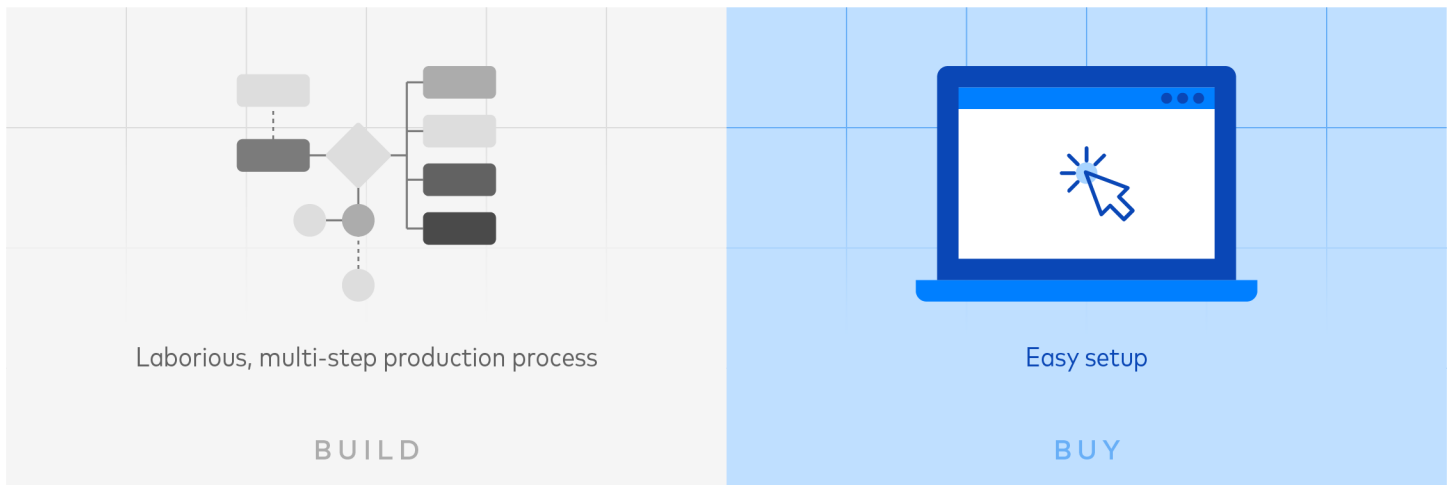
Beyond the time commitment, there is also the inherent complexity of building a reliable, performant piece of software. Building a data pipeline consists of the following steps:

1. Obtain developer access to the data source
2. Explore the data
3. Design the schema/data models
4. Set up a connector framework
5. Design an update and delete strategy
6. Test the connector and validate the data

Don't forget that a data pipeline must be updated whenever the underlying data source changes. The moment changes to the underlying data are detected, the process starts all over again. This maintenance cycle will continue indefinitely. It is possible, and more sensible, to outsource all pipeline building and maintenance to an outside vendor.

Continued, indefinite maintenance	No maintenance
BUILD	BUY

People (& Money)
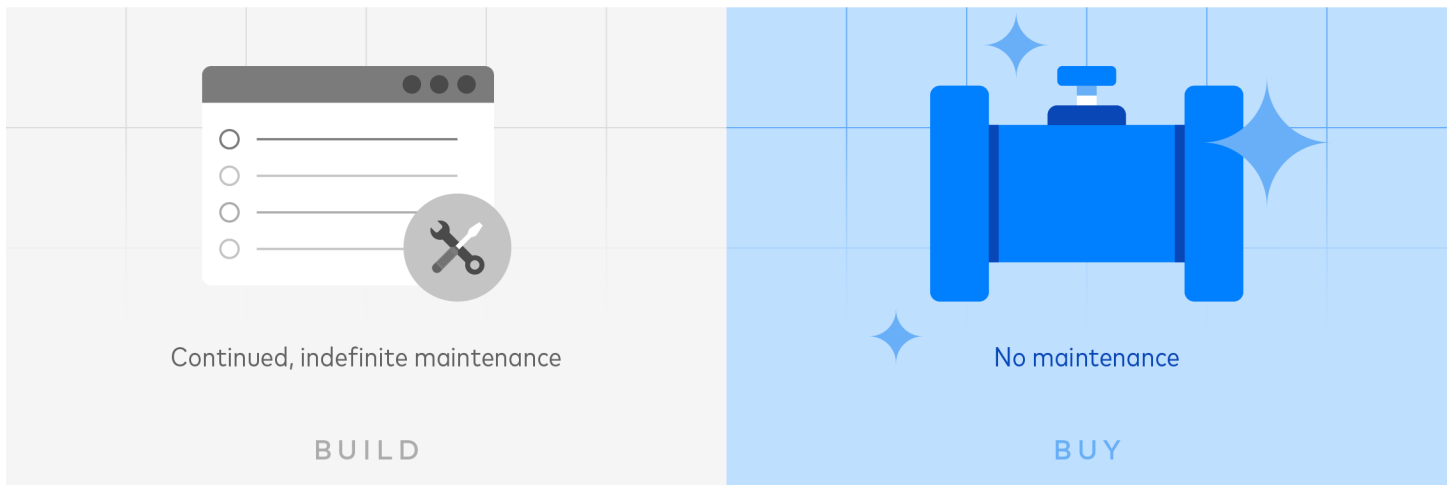
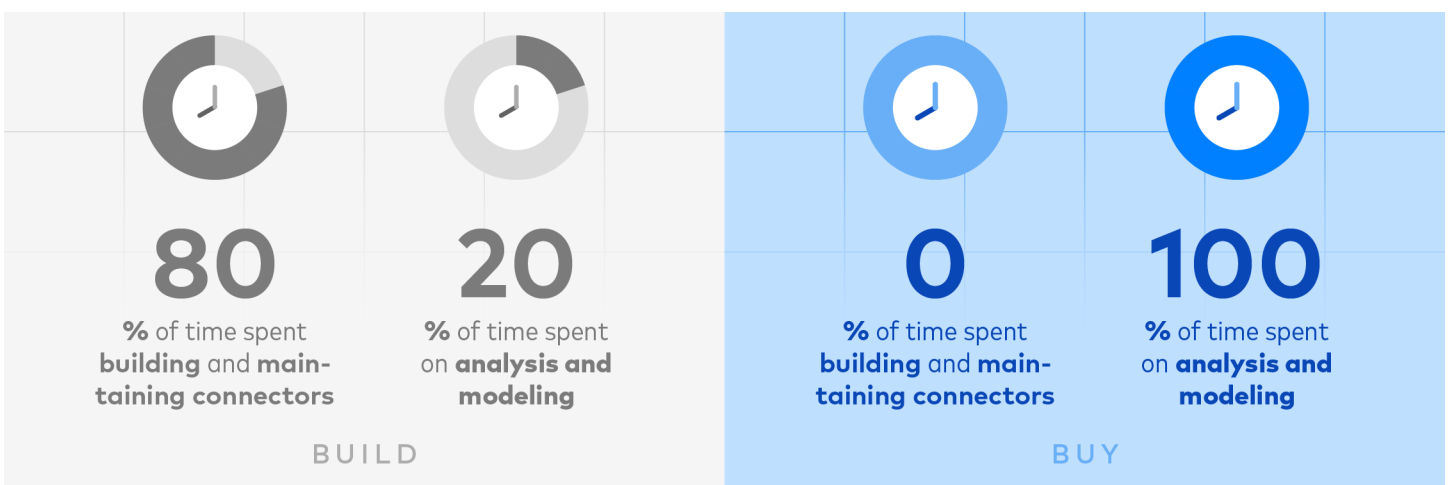|	
3 engineers to build and maintain	**0** humans, fully automated
BUILD	BUY

Based on our customers' experiences, a typical company requires the equivalent of at least two or three full-time data engineers to build and maintain a data pipeline. An automated, off-the-shelf solution makes human intervention unnecessary.

$500k
Cost of 3 full-time engineers

$50k
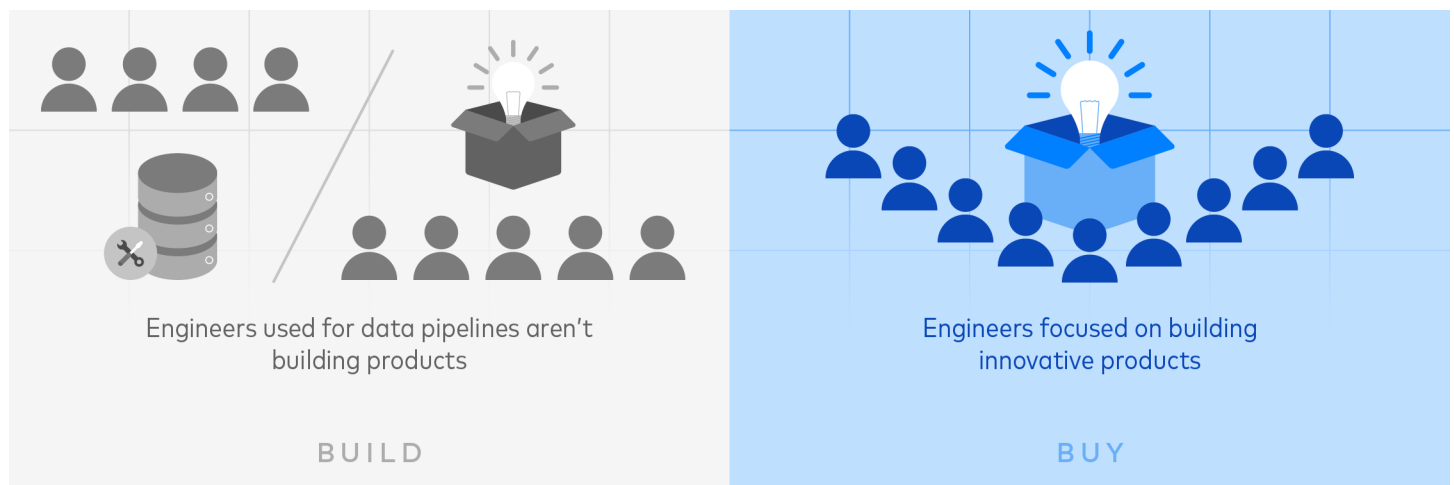Starting annnual subscription costs

BUILD

BUY
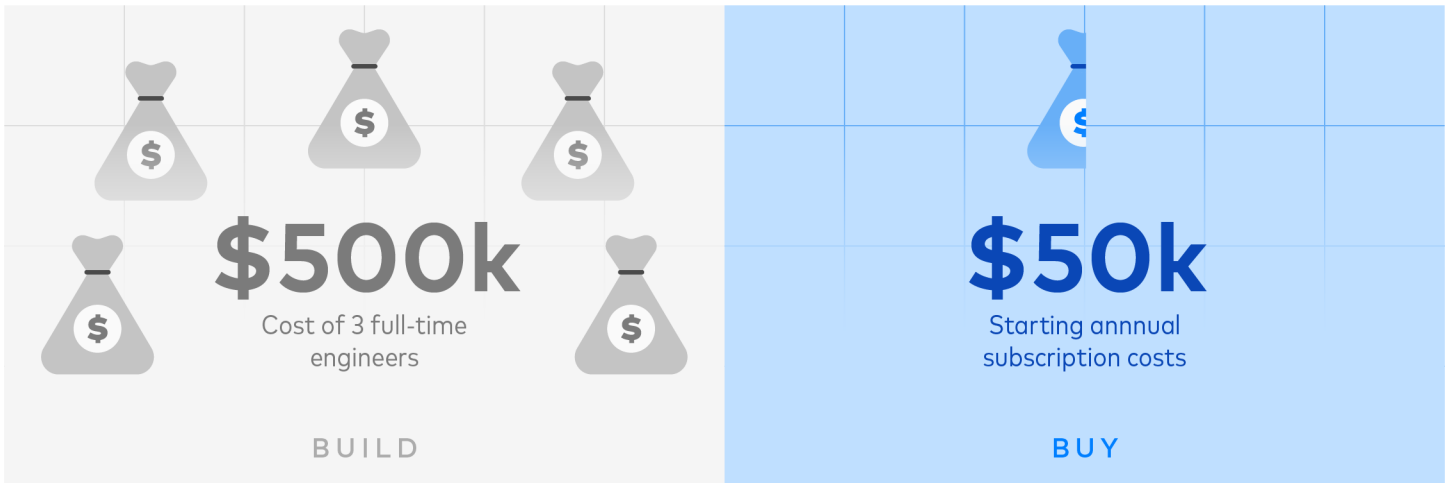
The total cost of three full-time engineers can reach the high six figures once benefits are accounted for. Depending on usage, subscriptions for automated data integration tools can start in the low five figures. The estimate of $50,000 per year is typical of a mid-sized company with five connectors.

Opportunity Costs

As a discipline, data engineering is not taught in any formal degree program and high-quality talent is scarce. Despite the difficulty of the practice, data integration duties often become the responsibilities of data scientists, analysts, and engineers.

80
% of time spent **building** and **maintaining connectors**

20
% of time spent on **analysis and modeling**

0
% of time spent **building** and **maintaining connectors**

100
**%** of time spent on **analysis and modeling**

BUILD

BUY

Job descriptions for data analysts and data scientists usually emphasize the importance of statistical methods, predictive modeling, and machine learning, yet data scientists famously spend about **80% of their time** on data integration rather than analysis. What if you could spend 0% of your time on data integration and 100% of it on machine learning, analytics, and other essential business activities?

Engineers used for data pipelines aren't building products

BUILD

Engineers focused on building innovative products

BUY

By the same token, engineers who are assigned to data integration activities have less time to spend building core products and performing other essential business functions.

Future Growth

As the size and obligations of your business continue to grow, and as new cloud-based tools continue to proliferate, you will likely continue to add additional data sources. The complexity and effort of building and maintaining a data pipeline for a huge number of data sources can quickly escalate beyond your data engineering team's ability to handle.

Increased complexity over time | Standardized schemas

BUILD | BUY
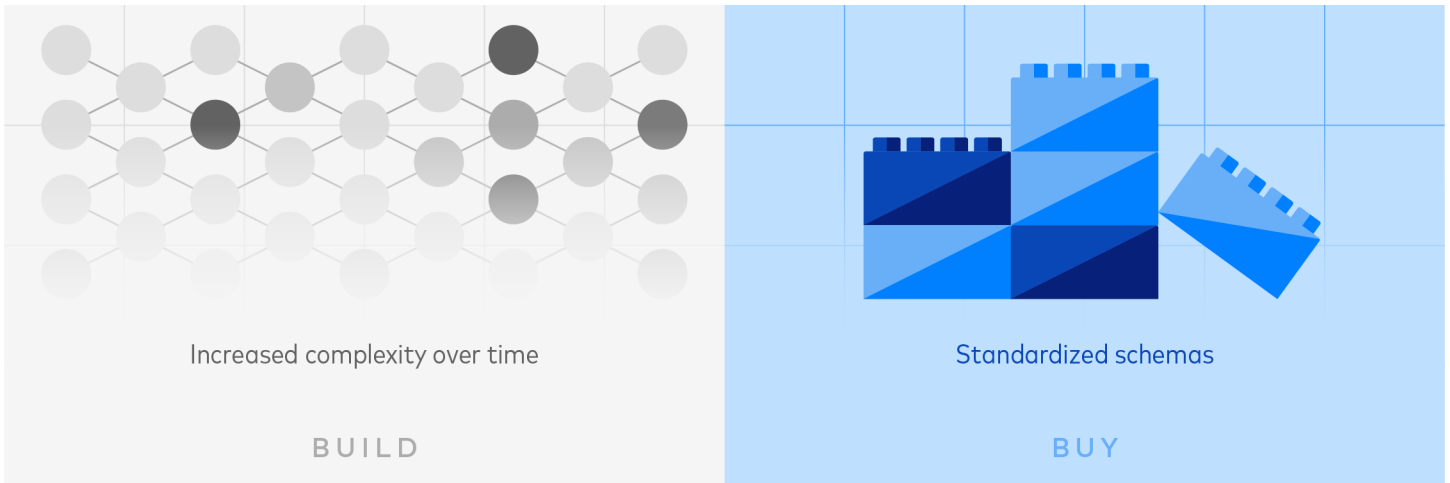
By contrast, if you use a data pipeline that features standardized schemas, you not only avoid the difficulty of building, maintaining, and reconciling data connections from a huge number of sources, but also leverage analyst templates and other derivative data products constructed from standardized schemas.

The Division of Labor Works for You

Performing data integration well involves a great deal of experience, trial, and error. Unless your core business is data engineering, there is little reason for you or your peers to develop the expertise to build data pipelines when off-the-shelf solutions exist, and there is even less reason to shunt these duties onto members of your team who lack the aptitude or desire to perform them.

Let someone who has already scaled the learning curve work for you, so you can spend your time, money, and energy building your core business and using data to guide it.