

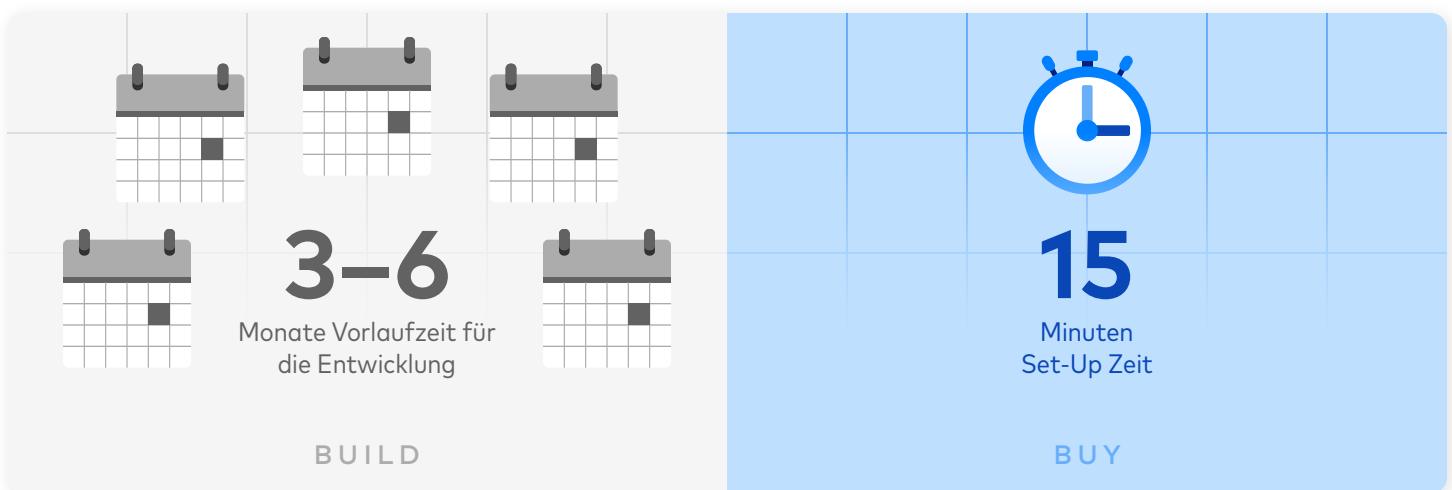
# Build vs. Buy : Ein illustrierter Leitfaden

Schauen Sie sich die folgenden  
Vergleiche an, bevor Sie  
anfangen, eine eigene Data  
Pipeline aufzubauen

# Build vs. Buy

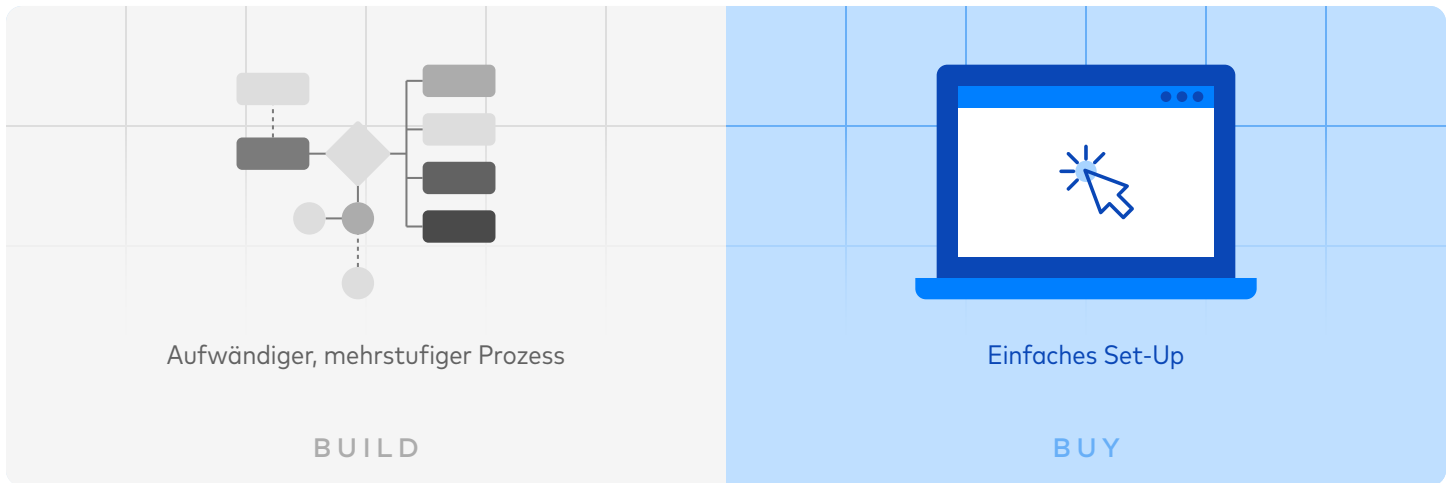
Ein durchschnittliches Unternehmen nutzt heute [weit über 100 Softwareanwendungen](#), von denen viele wertvolle Informationen über die Verfahren im Unternehmen enthalten. Ihr Unternehmen nutzt im Laufe der Zeit wahrscheinlich ähnlich viele Anwendungen, wenn nicht gar mehr. Daher benötigen Sie eine Lösung, mit der Sie die von den Anwendungen produzierten Daten integrieren können. Ziehen Sie auf ihrem Weg zur Datenintegration die Vorteile in Betracht, die Ihnen eine automatisierte, fertig erhältliche Lösung bieten kann. Werfen wir einen Blick auf die folgenden Vergleiche:

## Zeitlicher und finanzieller Aufwand



Das Erstellen Ihrer eigenen Pipeline erfordert einen bedeutenden Arbeitszeitaufwand. Nach den Erfahrungen unserer Kunden kann es drei bis sechs Monate dauern, eine grundlegende Pipeline einzurichten.

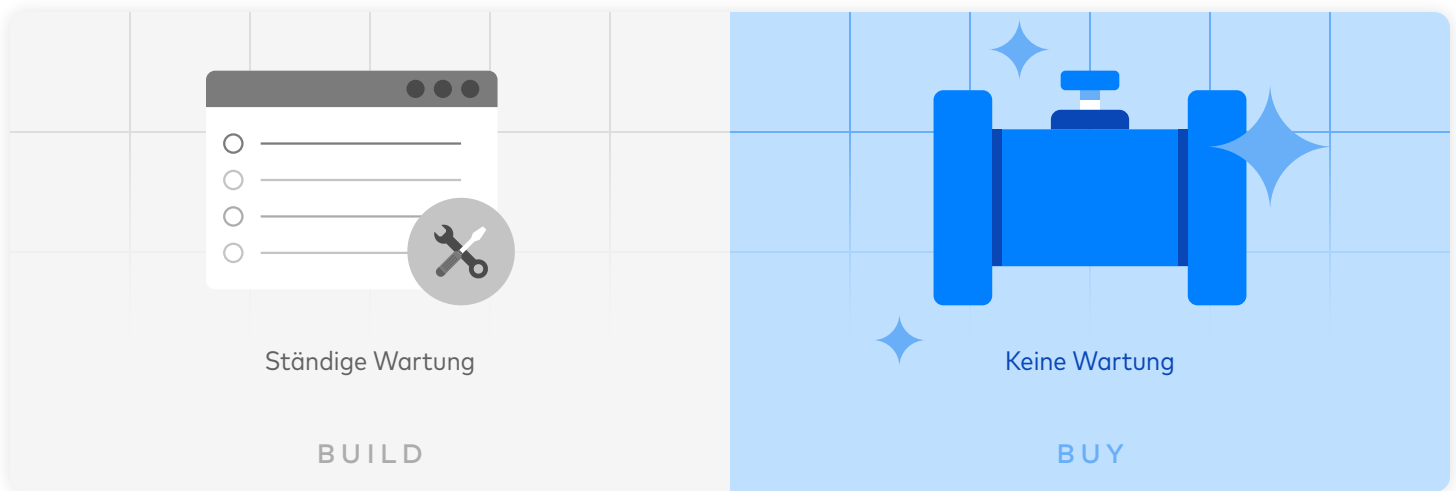
Eine handelsübliche Lösung mit vorgefertigten Konnektoren dagegen kann innerhalb weniger Minuten eingerichtet werden.



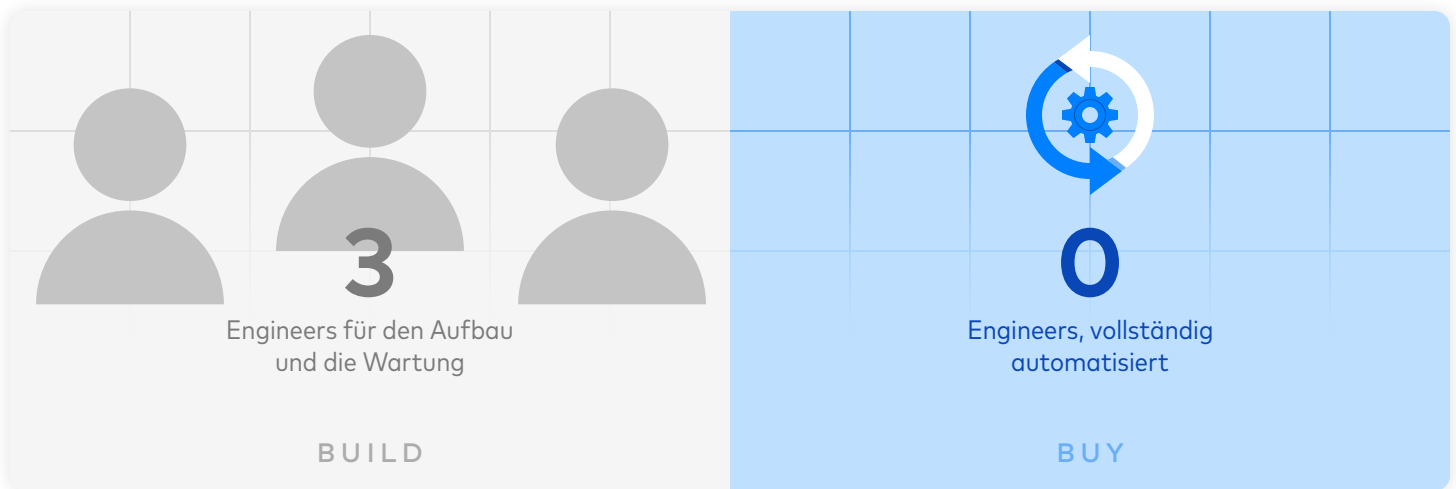
Neben dem Zeitaufwand spielt auch die Komplexität eine Rolle, die mit dem Erstellen einer zuverlässigen und leistungsfähigen Software verbunden ist. Die Erstellung einer Data Pipeline besteht aus folgenden Schritten:

1. Entwicklern Zugriff auf die Datenquelle verschaffen
2. Daten untersuchen
3. Schema-/Datenmodelle entwerfen
4. Konnektor-Framework erstellen
5. Aktualisierungs- und Lösungsstrategie erstellen
6. Konnektor testen und Daten validieren

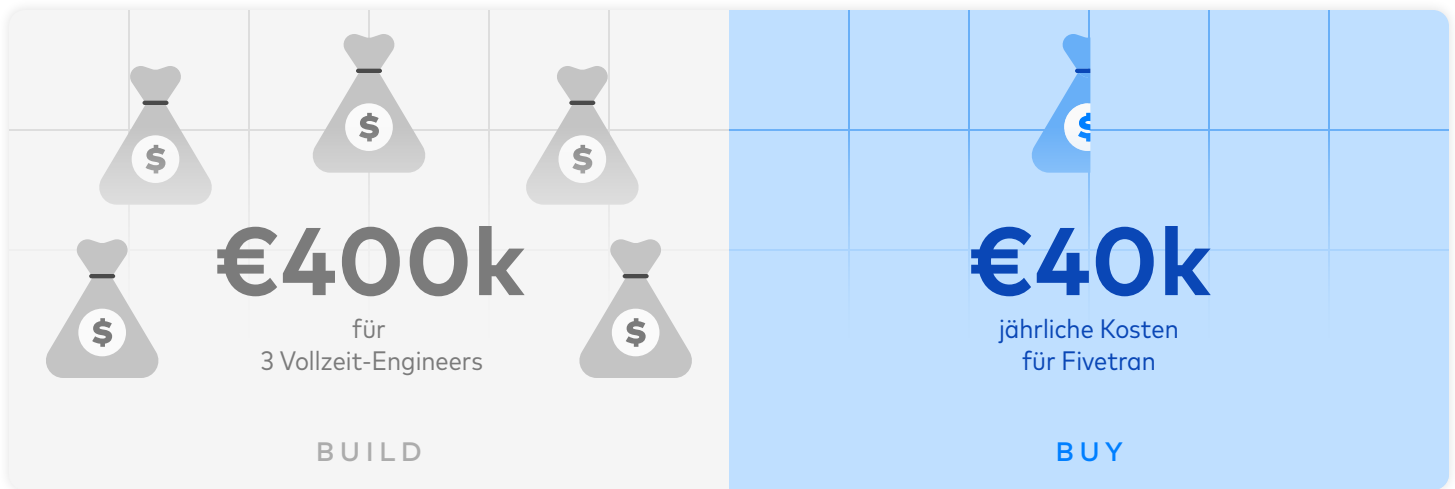
Denken Sie daran, dass eine Data Pipeline stets aktualisiert werden muss, sobald sich die zugrunde liegende Datenquelle ändert. In dem Moment, in dem Änderungen an den zugrunde liegenden Daten erkannt werden, beginnt der Prozess von vorn. Dieser Wartungszyklus wird sich immer wieder wiederholen. Es ist hingegen möglich und auch naheliegend, die gesamte Erstellung und Wartung der Pipeline einem externen Vendor zu übertragen.



## Personal- und Finanzaufwand



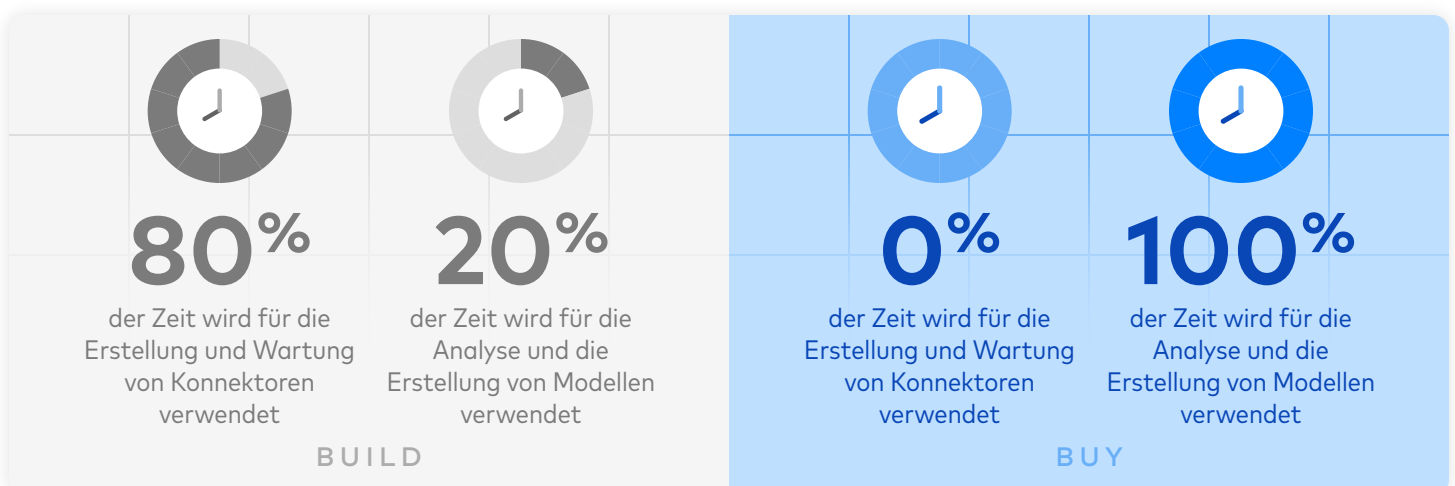
Nach den Erfahrungen unserer Kunden ist in einem typischen Unternehmen die Arbeitskraft von mindestens zwei oder drei Vollzeit-Dateningenieuren erforderlich, um eine Data Pipeline zu erstellen und zu warten. Durch eine automatisierte, fertig erhältliche Lösung werden manuelle Eingriffe überflüssig.



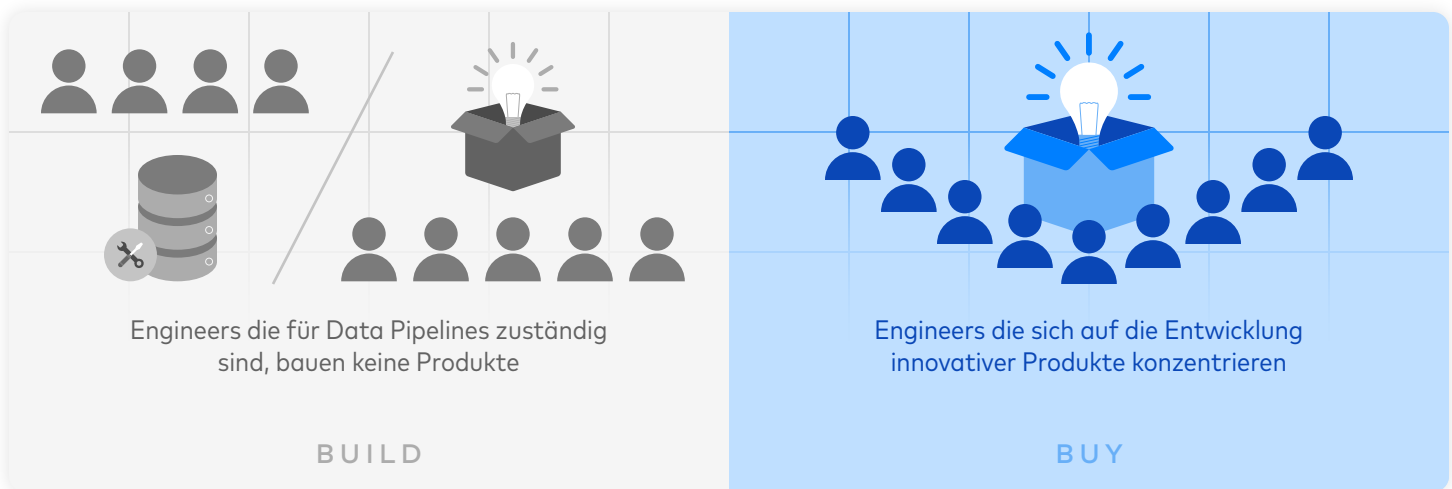
Die Gesamtkosten für drei Vollzeit-Dateningenieure können im oberen sechsstelligen Bereich liegen, wenn man die Arbeitsnebenkosten und Abgaben berücksichtigt. Je nach Nutzung sind Abonnements für automatisierte Datenintegrationstools bereits für niedrige fünfstelligen Beträge erhältlich. Für ein mittelständisches Unternehmen mit fünf Konnektoren wird der Aufwand normalerweise auf ungefähr 50.000 USD geschätzt.

## Opportunitätskosten

Data Engineering als Fach wird in keinem Studiengang formal gelehrt, und nicht zuletzt deshalb sind hochkarätige Fachkräfte in diesem Bereich rar. Trotz der Schwierigkeiten in der Praxis werden Aufgaben der Datenintegration oft Data Scientists, Analysten und Ingenieuren übertragen.



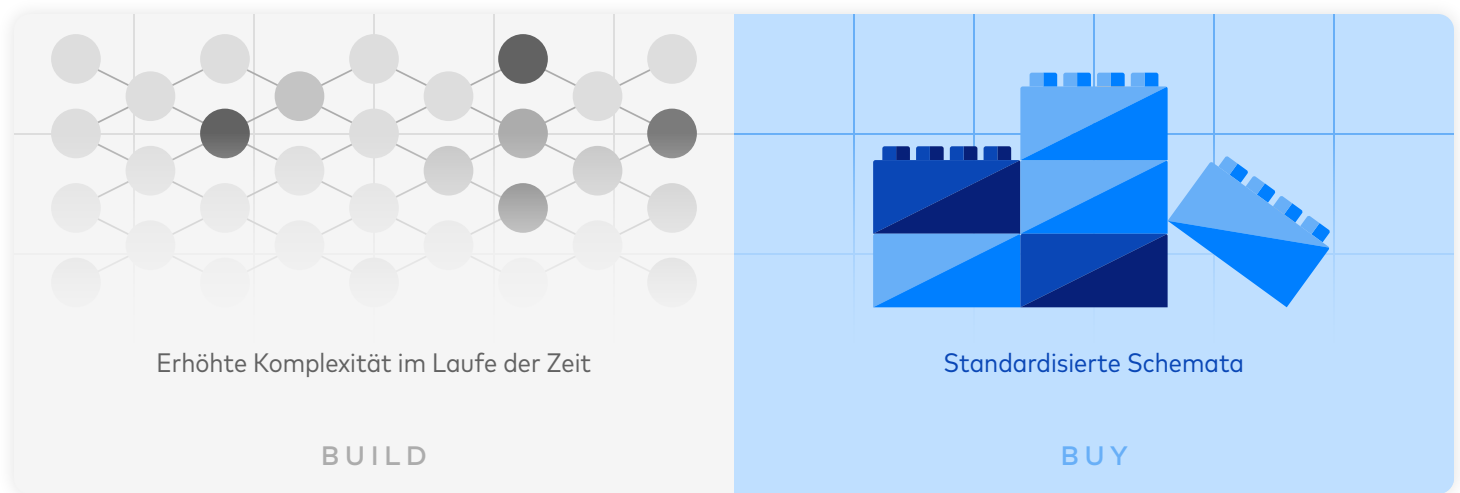
In Stellenbeschreibungen für Datenanalysten und Data Scientists wird normalerweise auf Kenntnisse zu statistischen Methoden, Vorhersagemodellen und maschinellem Lernen Wert gelegt. Bekanntermaßen verbringen Datenwissenschaftler jedoch etwa 80 % ihrer Zeit eher mit der Datenintegration als mit der Analyse. Wie wäre es, wenn Sie 0 % Ihrer Zeit für Datenintegration aufwenden müssten und stattdessen 100 % davon für maschinelles Lernen, Analysen und andere wichtige Geschäftsaktivitäten zur Verfügung hätten?



Umgekehrt haben Ingenieure, die mit der Datenintegration betraut sind, weniger Zeit zum Entwickeln von Kernanwendungen und für andere wichtige geschäftliche Funktionen zur Verfügung.

## Zukünftiges Wachstum

Da die Größe und die Aufgaben Ihres Unternehmens weiter zunehmen und sich neue cloudbasierte Tools verbreiten, werden Sie wahrscheinlich weitere Datenquellen hinzufügen. Komplexität und Aufwand bei Aufbau und Wartung einer Data Pipeline für eine große Anzahl von Datenquellen können schnell über die Fähigkeiten Ihres Data Engineering-Teams hinausgehen.



Wenn Sie dagegen eine Data Pipeline mit standardisierten Schemata verwenden, vermeiden Sie nicht nur die Herausforderungen beim Aufbauen, Verwalten und Abstimmen von Datenverbindungen aus einer Vielzahl von Quellen, sondern nutzen auch Analystenvorlagen und andere abgeleitete Datenprodukte, die aus standardisierten Schemata erstellt wurden.

## Profitieren Sie von der Arbeitsteilung

Eine gute Datenintegration erfordert viel Erfahrung, wozu auch viel Ausprobieren gehört. Sofern das Data Engineering nicht gerade Ihr Kerngeschäft ist und Standardlösungen verfügbar sind, gibt es für Sie oder Ihre Kollegen kaum einen Grund, sich das Fachwissen zum Aufbau von Data Pipelines anzueignen. Und es gibt noch weniger Gründe, diese Aufgaben auf Mitglieder Ihres Teams zu übertragen, denen die Fähigkeit oder Motivation dazu fehlen.

Nutzen Sie externes Expertenwissen, damit Sie Ihre Zeit, Ihr Geld und Ihre Energie darauf verwenden können, Ihr Kerngeschäft aufzubauen und von Daten leiten zu lassen.