

Modernisierung eines Data Stack für die effiziente Verwaltung komplexer Marketingmaßnahmen



"Die Marketing-Ausgaben sind einer unserer relevantesten Budget-Posten. Ungenauigkeiten in der Data Pipeline für's Marketing können zu kostenspieligen Fehlentscheidungen führen. Eine skalierbare, robuste und nachvollziehbare Data Pipeline spart signifikante Kosten."

Sascha Urban
Director Data
blinkist

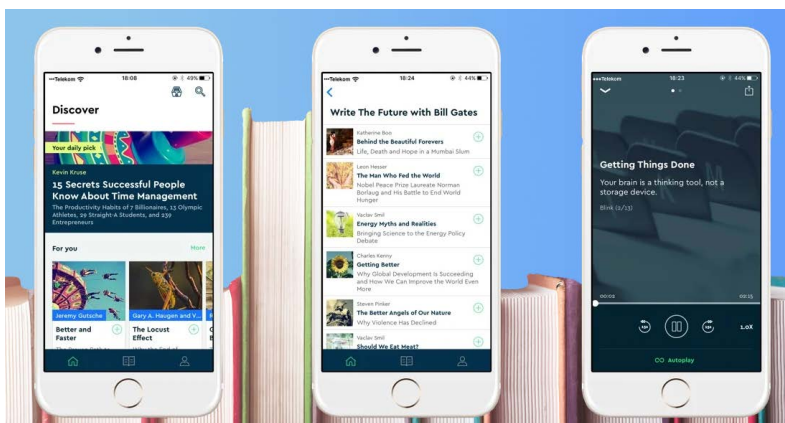


Aufgabe

Modernisierung eines veralteten Tech Stack für die fehlerfreie Verarbeitung und Analyse komplexer Daten.

Über HD+

Blinkist verwandelt die Kernaussagen der besten Sachbücher in clevere Kurztexte und Audiotitel in Hörbuchqualität. Blinkist ist eine App, die die großen Ideen der besten Sachbücher in einprägsame Kurztexte verpackt und erklärt. Die Inhalte der über 5.500 Titel starken Bibliothek reichen von Sachbuch-Klassikern über populäre Ratgeber bis hin zu diskutierten Neuerscheinungen. Basierend auf wissenschaftlichen Erkenntnissen wird jeder Titel von speziell geschulten Autoren und Autorinnen aufbereitet und den Usern als Kurztext und Audiotitel zur Verfügung gestellt. Blinkist verzeichnet weltweit über 20 Millionen Nutzende.



Projekt im Überblick



Kunde

Blinks Labs GmbH

Branche

Digitalindustrie

Lösung

Cloudbasierter Modern Data Stack und Data-Vault-Modellierung

Services

- > Modernisierung eines bestehenden Data Stack
- > Datenanbindung und Integration von Drittsystemen
- > Data-Vault-Modellierung
- > Technologisches Enablement

Technologien

- > Fivetran
- > Snowflake
- > dbt



Ausgangslage

Im Jahr 2012 gegründet, entwickelt sich Blinkist über die Jahre von einem erfolgversprechenden Start-up zu einem stark wachsenden Digitalunternehmen. Das liegt unter anderem an ausgeklügelten Marketingmaßnahmen, die sich zu einem großen und wichtigen strategischen Baustein entwickelt haben. Blinkist setzt auf eine ausgefeilte Multi-Channel-Strategie und bespielt zahlreiche unterschiedliche Kanäle mit dedizierten heterogenen Inhalten. Dazu gehören Plattformen wie Google Adwords, YouTube, Instagram, Microsoft News sowie unterschiedliche Content-Marketing-Angebote. Im Ergebnis stehen Blinkist äußerst umfangreiche Marketingdaten zur Verfügung, die konsolidiert und verarbeitet werden müssen, um stichhaltige Analysen vornehmen zu können. Bei gleichzeitig schnell steigenden Marketingausgaben ist es umso bedeutender, die Effizienz der gefahrenen Kampagnen berechnen zu können.

Stark wachsendes Datenvorkommen

Doch Blinkist stehen nicht nur umfangreiche Daten aus dem Marketingbereich zur Verfügung. Auch in den Bereichen Sales sowie Customer Management fallen enorme Datenmengen an. Dazu gehören Informationen über rund 20 Millionen Abonnenten und Abonnentinnen sowie Probe-Abonnements weltweit, deren Vorlieben und Nutzungsverhalten oder die Verarbeitung von verschiedenen Abrechnungsvarianten wie Paypal, Kreditkarten, App Stores wie Google und Apple und die damit zusammenhängenden Versteuerungsoptionen.

Konnte Blinkist anfangs auf eine kleine Datenbasis zugreifen, die Inhouse und mit soliden BI-Kenntnissen verwaltet wurde, werden in neuerer Zeit stark wachsende Datenvolumen verzeichnet. Das Data Management gestaltet sich somit technisch zu anspruchsvoll für bisherige Lösungen, was vor allem auch an einem nicht mehr zeitgemäßen Tech Stack liegt. So schleichen sich teils erhebliche Fehler in der Datenauswertung ein, die vor allem technisch begründet sind.

Fehleranfällige technische Infrastruktur

Gravierende Komplikationen finden sich vor allem bei Google-Kampagnen, deren Planungen sich auf penibel

vorbereitete Datenauswertungen stützen. Leider entpuppen sich die verwendeten Daten als falsch, da die Zuordnung der Datenquellen (Regionen und Länder) und die technische Verarbeitung der Daten mangelhaft sind. Hierdurch entsteht ein markanter finanzieller Schaden, der offensichtlich in einer fehleranfälligen technischen Infrastruktur begründet liegt.

Somit ist der Zeitpunkt gekommen, an dem sich das Data Team von Blinkist professionelle externe Unterstützung holt, um zwei grundlegende Probleme zu lösen:

1. Wie kann die Verarbeitung und Analyse äußerst komplexer heterogener Daten aus Marketing, CRM und Sales fehlerfrei gelingen?

2. Welche Elemente des Tech Stack bedürfen einer Modernisierung?



Ziele

Im Vordergrund steht die Modernisierung der technischen Infrastruktur. Die bestehenden Module sollen überprüft und hinterfragt werden, um die effiziente und zuverlässige Datenanalyse zu gewährleisten. Gebrochene Data Pipelines und unbemerkte Fehler in der Datenverarbeitung sollen der Vergangenheit angehören. Das setzt eine enge Zusammenarbeit mit dem Data Team von Blinkist voraus, dessen Enablement eine Kernaufgabe von taod ist.

Projektverlauf

Ausgangspunkt für die Überarbeitung der bestehenden Datenlandschaft ist ein Proof of Concept. Als Use Case dient die Erhebung der sogenannten Country-Distribution: In welchem Land investiert Blinkist wieviel Budget? Da Blinkist zu diesem Zeitpunkt im Tagesgeschäft hauptsächlich mit der Fehlerbehebung innerhalb des Tech Stack beschäftigt ist, sind keine Kapazitäten mehr für weitreichende und grundlegende Datenanalysen vorhanden. Die aktuelle Datenbasis ist zu komplex, so dass sich die inhaltliche Auseinandersetzung mit den gewonnenen Informationen als zu anspruchsvoll herausstellt. Als klare Zielsetzung des PoC steht der Aufbau neuer Data Pipelines. Zudem sollen die Fehlerquellen eruiert werden, die zu den massiven Verfälschungen der Analyseprojekte führen.



Jetzt Beratung anfragen!



Proof of Concept als Game Changer

Ein Proof of Concept dient einer pragmatischen und fokussierten Projektvorgehensweise. Bevor ein größer dimensioniertes Datenprojekt gestartet wird, ermöglicht der Proof of Concept die Validierung des Vorhabens im Rahmen eines kleineren "Vorprojekts". Der direkte Projektbezug erleichtert die Bewertung der geplanten Maßnahmen und liefert sowohl schnelle als auch aussagekräftige Ergebnisse. Das weitere Vorgehen profitiert von den gewonnenen Erkenntnissen und kann dementsprechend granularer geplant werden.

Data-Use-Case-Workshop für konkrete Ergebnisse

In einem Data-Use-Case-Workshop entwickelt taod gemeinsam mit dem Finance-Team von Blinkist ein Proof of Concept. Die Marketingausgaben sind enorm wichtig für Blinkist. Mehr als die Hälfte des Budgets wird in Marketing investiert oder stellt eine notwendige Investition für effektives Marketing dar. Durch inkorrekte Datentransformationen entstehen allerdings massive sowie teils unerkannte Fehler, die sich über die gesamte Data Journey hinwegziehen. Der ROI des Budgets ist somit teilweise fehlerhaft, was wiederum zu Fehlallokationen im Budget führt. Auch wenn dies nur in etwa einem von zwanzig Fällen vorkommt, entsteht dennoch relevanter Schaden.

Schlussendlich muss deshalb eine klare Zuordnung der Investitionen erfolgen, damit ersichtlich wird, wo genau welche Gelder ausgegeben werden. Bestehende Unstimmigkeiten zwischen selbst erstelltem Reporting und tatsächlicher Abrechnung müssen beseitigt werden. Dieser Use Case soll Basis für die weitere Optimierung der Analyseaktivitäten werden.

Das Blinkist Data Team wünscht sich granulare Auswertungsmöglichkeiten, um einen detaillierten Blick für jeden einzelnen Abonnenten und jede einzelne Abonnen-



tin zu entwickeln. Im Laufe des Workshops wird schnell deutlich, dass der Data Stack von Blinkist grundlegend erneuert werden muss. Insbesondere auch deshalb, weil sich das Bug Fixing mitunter über Tage hinweg zieht und keine Zeit mehr für die wertschöpfende Arbeit mit Daten bleibt. Der ursprüngliche Gedanke, jede einzelne Fehlerquelle ausfindig zu machen und zu korrigieren wird zugunsten einer produktiven Arbeitsweise schnell verworfen. Der Aufwand, die bestehende Infrastruktur anzupassen ist ungleich höher, als sie durch einen Modern Data Stack zu ersetzen. Im ersten Zug werden jedoch nicht sämtliche Data Pipelines ersetzt. Blinkist setzt auf eine sukzessive Umstellung, um den Arbeitsaufwand über einen längeren Zeitraum zu entzerren.

Hoher Detailgrad und lückenlose Datenverfügbarkeit

Der bisherige Tech Stack setzt sich aus Amazon Redshift als Data Warehouse, Matillion für die Transformation der Daten sowie Periscope Data für die Datenanalyse zusammen. Eine hohe Fehleranfälligkeit wird in den bestehenden Data Pipelines ausgemacht, die völlig neu aufgebaut werden müssen. Viele Quellen von Blinkist, wie beispielsweise Google, bereiten Daten bereits auf und liefern fertige Reports. Solche Reports sind für kleine Unternehmen sehr hilfreich, da sie auf die Expertise in der Erstellung solcher Reports verzichten können. Je spezifischer aber die Anforderungen werden, desto schwieriger wird es auch, konkrete Fragen anhand solcher Reports zu beantworten. Auch kommt es vor, dass die besonders ausführlichen Reports beabsichtigt unvollständig sind, weil sie sonst zu umfangreich für die meisten Nutzenden wären.

Um eine zuverlässige Datengrundlage für verschiedenste Analysen zu schaffen, ist es für Blinkist notwendig, einen hohen Detailgrad und eine lückenlose Datenverfügbarkeit zu schaffen. Einzelne Reports automatisiert auf fehlende Daten hin zu prüfen und diese Mängel beheben zu lassen, gehört zu den wichtigsten und schwierigsten Aufgaben eines Data Teams. Diese Prüfungen und Qualitätsverbesserungen sollen mit neuen Werkzeugen umgesetzt werden.

Intelligente Werkzeuge kombinieren Flexibilität und Perfektion

Die Wechsel des Konnektoren-Tools von Data Virtuality über versuchsweise Airbyte hin zu Fivetran, um vor allem interne Ressourcen zu entlasten, sowie des Transformations-Tools von Matillion zu dbt sorgen künftig für enorme Flexibilität und lassen eventuelle Anpassungen rasch und unkompliziert zu. Das ist elementar, denn Blinkist liefert ein digitales Produkt, das nahezu endlos skaliert werden kann. Bis dato besteht bereits eine große Medienbibliothek mit mehr als 5.500 Titeln. Dabei entstehen Kosten für Zahlungsabwicklung und Cloud Services, die wiederum mit der Anzahl der zahlenden User skalieren.



Die Anforderungen an Zuverlässigkeit, Verfügbarkeit und Korrektheit von marketingrelevanten Daten wie Ausgaben und die dadurch generierten Einnahmen sind also enorm hoch. Zudem ist Blinkist ein junges und stark wachsendes Unternehmen, das sich mit stetig wandelnden Prozessen auseinandersetzt. Demzufolge ist eine flexible technische Infrastruktur elementar für sämtliche businessgetriebenen Aktivitäten.

Hier entsteht ein Zielkonflikt: Zuverlässigkeit und Korrektheit erfordern langsames, sauberes Arbeiten, ausgiebiges Testen und möglichst wenige Änderungen, sobald ein hoher Perfektionsgrad erreicht ist. Flexibilität erfordert das genaue Gegenteil: Schnelligkeit, Spontaneität und Mut zu Fehlern.

Deshalb sind zwei Schritte wesentlich:

1. Werkzeuge wählen, die zur Lösung dieses Zielkonflikts geschaffen sind.

2. Data Vault als modernen Ansatz der Datenmodellierung wählen

Als Transformationswerkzeug senkt dbt den Aufwand für gründliches Testing um bis zu 80 Prozent. Datenmodellierung beschreibt die Strukturierung von Daten in miteinander verknüpften Tabellen. Das Data-Vault-Model hat einen klaren Fokus: Flexibilität bei Einhaltung von hohen Qualitätsstandards. Bei einem Data Vault Model können bildlich gesprochen beinahe gleichzeitig das Fundament, die Mauern und das Dach gebaut und einfach nachträglich ergänzt werden. Dadurch lassen sich hohes Tempo und langfristig nutzbare Lösungen perfekt miteinander kombinieren.

Datenmodellierung und Entwicklung von Logiken via Data Vault

Neben der gesamten Überarbeitung der Datentransformation und dem Wechsel des Konnektoren-Tools entwirft taod neue Logiken, um den dezidierten Analyseanforderungen gerecht zu werden. Zum Einsatz kommt die Data-Vault-Methode, die sich besonders für agile Data Warehouses eignet und die flexible Verarbeitung komplexer Datensätze sichert.

Ein Beispiel hierfür sind zwei zentrale Google Ads Reports von Blinkist. Report 1 umfasst 100 % der Marketingausgaben, bietet jedoch 0 % geographische Informationen. Report 2 umfasst 80 % der Marketingausgaben, zeigt für diese 80 % jedoch auch die gesamten Geo-Daten an. Blinkist möchte einen Report, der sämtliche Ausgaben mit sämtlichen Geo-Daten zu 100 % abbildet.

Künftig wird daher für jede einzelne Anzeige stets tagessaktuell berechnet, wie groß die Differenz zwischen den Ausgaben in den beiden Reports ist. Die Differenz wird dann als neuer Eintrag mit der Geo-Information "weltweit" ergänzt. Auch wenn diese Lösung naheliegend wirkt, ist sie sehr wichtig und konnte mit der alten Lösung nicht umgesetzt werden. Dort wurde die Geo-Zuordnung nur geschätzt.

Die alten Informationen basierten entweder auf falschen Zahlen oder unzuverlässigen Geo-Daten. Die neuen Informationen ermöglichen die Abbildung der korrekten Summe für jeden Tag und jede Werbeanzeige mit maximalem Detailgrad in den Geo-Daten. Google selbst liefert nicht für jede Anzeige die Geo-Daten, weil User beispielsweise geschützte Netzwerkverbindungen wie VPN einsetzen.

Was ist Data Vault?

Die von Dan Linstedt geschaffene Methode Data Vault sorgt für die schnelle sowie korrekte Integration von Daten in ein Data Warehouse. Auf große inhaltliche Änderungen kann flexibel reagiert werden. Es handelt sich um eine Modellierungstechnik für insbesondere agile Data Warehouses.

Bei der Data-Vault-Modellierung werden Daten nicht wie bei gängigen Methoden in der 3. Normalform abgelegt, sondern alle zu einem Geschäftskonzept gehörende Informationen in drei Kategorien eingeteilt und in drei Typen von Datenbanktabellen abgelegt. Die strikte Trennung zwischen sogenannten Hubs, Links und Satellites ist wesentlich für die Integration von Daten aus mehreren Quellen, die Erweiterbarkeit des Datenmodells und komplexe ETL-Prozesse.

Aufbau eines Modern Data Stack mit Snowflake und dbt

Das bestehende Data Warehouse Amazon Redshift wird durch das Cloud Data Warehouse Snowflake ersetzt, das künftig für eine hohe Skalierbarkeit nach Bedarf sorgt. Der Aufbau des Data Stack wird zunächst von taod geführt und anschließend begleitet und sorgt auf diese Weise für ein hohes Enablement von Blinkist.

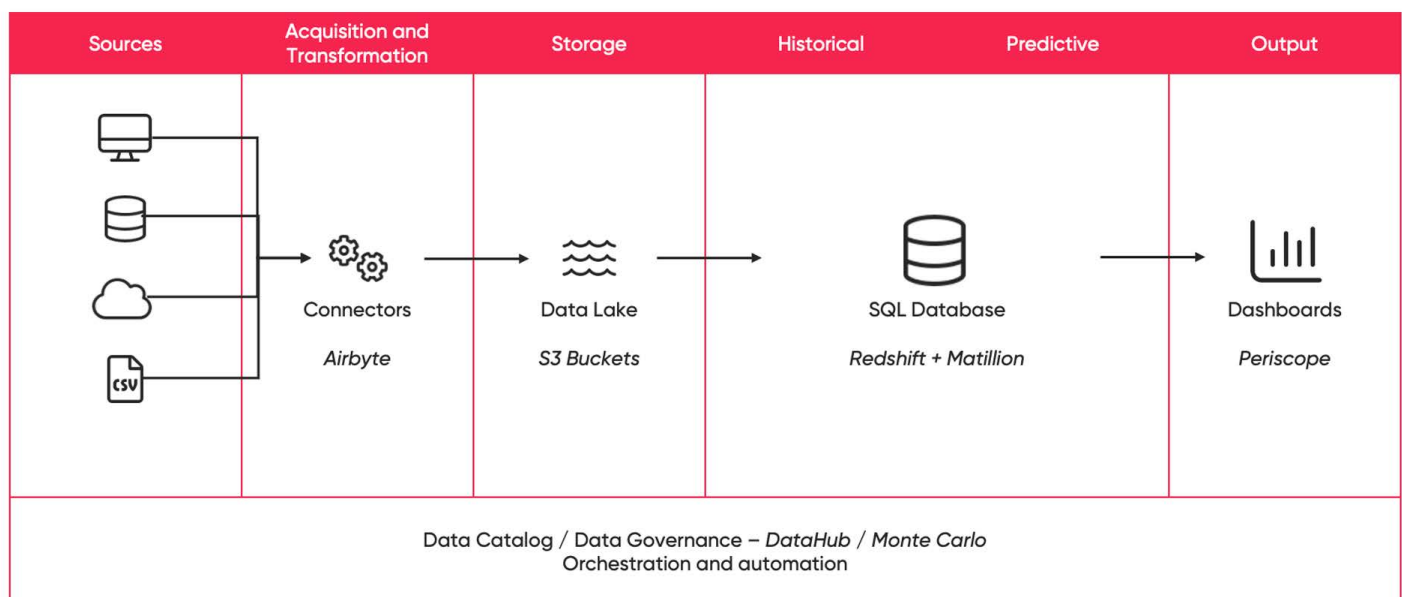
Data Vault stellt zudem eine neue Modellierungstechnik für das Data Team von Blinkist dar, die durch 1:1 Coachings im Laufe des Projekts geschult wird. Zudem reduziert sich durch Technologie und Enablement die Zeit für das Bug Fixing. Dauerten Korrekturen bislang schlimmstenfalls Tage, können Fehler nun innerhalb weniger Minuten erkannt und behoben werden.

Nach einer Projektlaufzeit von rund vier Monaten verwalten die Data Engineers, Data Scientists und Data Analysts von Blinkist den neuen Tech Stack eigenständig. Ab diesem Zeitpunkt übernimmt taod eine beratende Rolle und kümmert sich in regelmäßigen Intervallen im Rahmen eines FAQ-Formats um die inhaltliche Fortführung des Projekts.

Herausforderungen

- > Kombination aus alten und neuen Bestandteilen des Modern Data Stack
- > Verortung und Eliminierung der Fehlerquellen
- > Datenmodellierung und Aufbau neuer Logiken für komplexe Daten

Alter Tech Stack von Blinkist auf Basis von Amazon Redshift und Matillion



“Durch die Wahl der richtigen Werkzeuge konnten wir **Lösungen bauen, die höchsten Qualitätsstandards gerecht werden** und gleichzeitig an das Data Team von Blinkist zur eigenen Wartung übergeben werden konnten. Hätten wir die **bisher verwendeten Technologien weiterverwendet**, wäre der Aufwand vermutlich viermal so hoch gewesen und **hätte eine starke Abhängigkeit** von unserer Hilfestellung zur Folge gehabt. Es freut mich, dass wir Blinkist mit dem **Einsatz modernster Werkzeuge der Datenbereitstellung** enablen und damit zukunftsfähig machen konnten.”



Andreas Huppert
Managing Consultant
taod Consulting

Ergebnis: Modernisierter Tech Stack und hohes Enablement

Der aktualisierte Modern Data Stack sorgt technisch für die reibungslose Verarbeitung der komplexen Daten. Die hohe Fehleranfälligkeit konnte auf ein Minimum reduziert werden, auch deshalb, weil die neue Modellierungs-Technologie in Kombination mit modernen ETL-Prozessen Fehlerquellen von vornherein deutlich reduziert. Sämtliche bekannte Use Cases, die besonders die Marketingmaßnahmen des Unternehmens betreffen, sind mit dem neuen Tech Stack zuverlässig umsetzbar. Das Data Team von Blinkist wird während der gesamten Projektlaufzeit aktiv in den Umbau involviert. Es eignet sich durch begleitende Schulungsmaßnahmen sowie die enge Zusammenarbeit mit den taod Experts neue und vertiefende technologische Kompetenzen an.

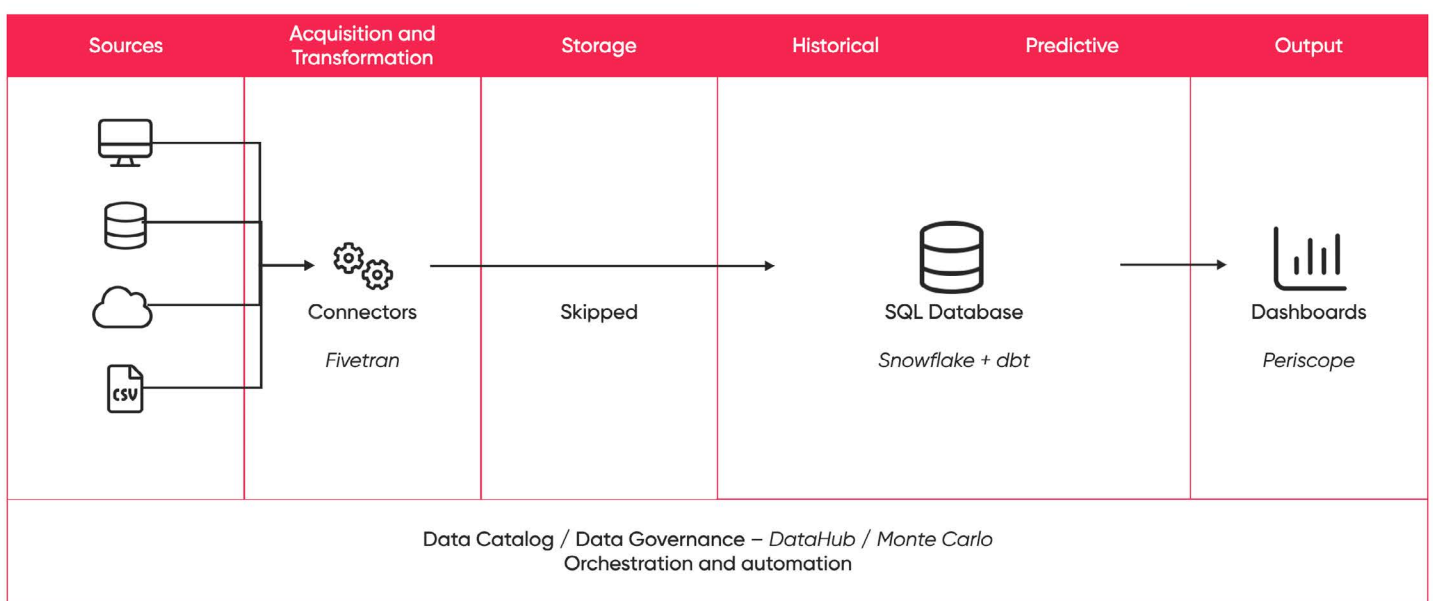


- > Proof of Concept und Use-Case-Workshop
- > Consulting für Data Architecture
- > Cloud Data Warehouse
- > Data Pipelines
- > Projektbegleitendes Enablement



Jetzt Beratung anfragen!

Neuer Modern Data Stack für Blinkist auf Basis von Snowflake und dbt





Wie können wir dich beraten?

Kontaktiere uns gerne unverbindlich, wenn du dich für das Thema Modern Data Stack interessierst und dich von uns beraten lassen möchtest.

Andreas Huppert
Managing Consultant
+49 151 53429328
andreas.huppert@taod.de



Jetzt Beratung anfragen!

Raise the potential in every byte

Hinweis: Zur besseren Lesbarkeit wird in dieser Case Study, neben Doppelformen und Partizipialformen, das generische Maskulinum verwendet. Die verwendeten Personenbezeichnungen beziehen sich – sofern nicht anders kenntlich gemacht – auf alle Gender.

Verwendete Bilder

taod Consulting GmbH
[istockphoto.com/Wirestock](https://www.istockphoto.com/Wirestock)
[istockphoto.com/artisteer](https://www.istockphoto.com/artisteer)
[istockphoto.com/cyano66](https://www.istockphoto.com/cyano66)
[istockphoto.com/kot63](https://www.istockphoto.com/kot63)
[istockphoto.com/Tatabrada](https://www.istockphoto.com/Tatabrada)

Datenschutzbeauftragter

Frank Gundlach
GTB – Genossenschafts-Treuhand
Bayern GmbH Wirtschaftsprüfungsgesellschaft

Türkenstrasse 22 - 24
80333 München
+49 170 9416034
fgundlach@gv-bayern.de

Kontakt

taod Consulting GmbH
Oskar-Jäger-Str. 173, K4
50825 Köln

+49 221 975 849 70
info@taod.de

Vertreten durch

Simon Biela, Matthias Steinforth,
Benedikt Stienen
Amtsgericht Köln HRB 95089

