

alteryx

Self-Service Analytics auf der Google Cloud Plattform:

Fünf Lektionen zur Datenvorbereitung,
die den Erfolg sichern



Sie möchten die Herausforderungen der Unternehmensanalyse selbstständig bewältigen und produktiver und innovativer werden. Doch dafür benötigen Sie Daten.

In der heutigen dynamischen und komplexen Datenlandschaft sind herkömmliche Data-Warehouse- und Business-Intelligence-Lösungen nicht mehr ausreichend. Selbst wenn Sie herkömmliche Data Warehouses mit Lösungen wie Tableau oder Qlik erschließen, arbeiten Sie letzten Endes immer noch in Tabellen und benötigen IT-Support.

Die Antwort ist eine Self-Service Analytics-Lösung.

Dieses Whitepaper beschreibt fünf Lektionen, die Alteryx daraus gezogen hat, Unternehmen wie das Ihre bei der Einführung von Self-Service Analytics zu unterstützen. Es konzentriert sich auf die Google Cloud Platform, eine umfassende intelligente Analytics-Suite, die Data Analysts die Freiheit gibt, mit jeder Form von Daten zu arbeiten, unabhängig von Größe und Geschwindigkeit.





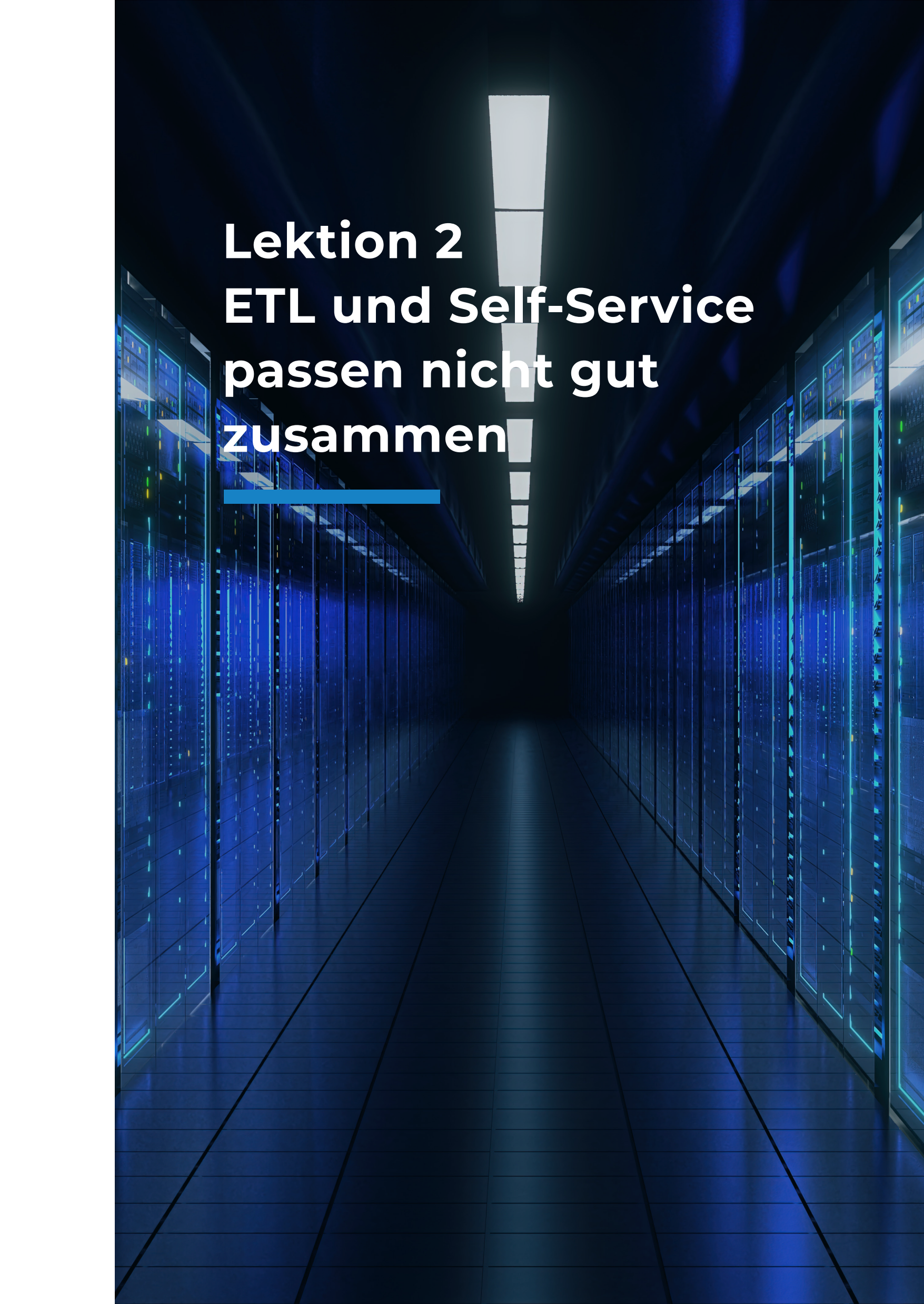
Lektion 1 Schluss mit Lift-and-Shift in der Cloud

Wenn Analytics in die Cloud verlagert wird, verfolgen viele Unternehmen einen Lift-and-Shift-Ansatz: Analyse-Anwendungen werden auf einem virtuellen Rechner in der Cloud installiert. Dieser Ansatz erleichtert zwar die Wartung der Infrastruktur, verbessert jedoch nicht die Art und Weise, wie Analysen bereitgestellt werden, bietet keinen großen Mehrwert für Ihr Geschäft und senkt nicht die Betriebskosten. Prozesse und Tools bleiben gleich und auch die einzelnen Benutzer:innen erleben keinen echten Mehrwert, wenn es um Self-Service geht.

Um Self-Service wirklich zu verwirklichen, müssen Analyse-Lösungen Cloud-nativ sein. Prozesse für die Einspeisung, Speicherung und Verarbeitung von Daten sowie die Berichterstattung müssen so gestaltet sein, dass sie nativ in Systeme integriert werden können, die speziell für die Cloud entwickelt wurden.

Sie müssen Cloud-Umgebungen unterstützen, die dynamisch, flexibel, skalierbar, zunehmend containerisiert und auf die Bereitstellung von Microservices ausgerichtet sind. Anstatt einfach die bisherigen Datenplattformen zu kopieren, die sie lokal vor Ort hatten, setzen erfolgreiche Unternehmen auf Cloud-Flexibilität und verwenden Open-Source-Systeme wie Kubernetes oder Docker, um Container und Datenflüsse zu orchestrieren.

Die Google Cloud Platform ist für Self-Service Analytics attraktiv, da sie eine native, serverlose Smart-Analytics-Suite bietet. Jede Komponente lässt sich einfach aktivieren oder dynamisch skalieren und die Ressourcenzuteilung wird je nach Nutzung verringert. Wartungsarbeiten müssen nicht geplant werden. Mit der Freiheit, jede Analysekomponente zu nutzen, Ressourcen flexibel einzusetzen und Kosten zu kontrollieren, können sich Unternehmen auf Daten und den Mehrwert konzentrieren, den sie für das Unternehmen bieten.



Lektion 2 ETL und Self-Service passen nicht gut zusammen

Ältere Datenintegrationstechnologien, wie Extract-Transform-Load (ETL), können verwendet werden, um Pipelines zu erstellen und Daten, die lokal und in anderen Clouds gespeichert sind, auf die Google Cloud Platform zu verschieben. Alternativ können neue Cloud-spezifische Datenintegrations- und Streaming-Technologien wie StreamSets, Fivetran oder Google Cloud Data Fusion effektiv Daten zwischen Systemen übertragen und für Analysen in der Google Cloud Platform verfügbar machen.

Es mag verlockend sein, ETL auch nach der Befüllung eines Data Lakes oder Data Warehouses weiter zu verwenden. Aber ETL ist einfach nicht für Self-Service Analytics geeignet.

Es ist eine ziemlich technische und komplexe Technologie, die sich an Dateningenieuren richtet, die mit klar definierten, wiederkehrenden Echtzeit- und Batch-Umwandlungen für umfangreiche Datenbewegungen arbeiten. Damit Fachleute, die sich mit Unternehmensdaten auskennen, auf die Daten zugreifen und sie für ihre Analysen nutzen können, ist zusätzliche Datenvorbereitung erforderlich.

Als Teil der Smart Analytics Suite bietet die Google Cloud Platform eine Datenvorbereitungslösung: Cloud Dataprep by Trifacta. Sie bewertet die Datenqualität, verfeinert, standardisiert, bereinigt und kombiniert Daten und führt verschiedene Datenberechnungen durch.

Cloud Dataprep by Trifacta wurde entwickelt, um eine ganze Reihe von Datenfachleuten – Dateningenieure, Datenanalysten, Geschäftsanalysten und andere datenorientierte Fachkräfte – zu unterstützen, einen Data Lake oder ein Data Warehouse zu erschließen, indem sie mit den Inhalten der Daten interagieren, um sie iterativ zu verfeinern und zu integrieren und nachgelagerte geschäftsorientierte Analysen zu ermöglichen.

Lektion 3 Self-Service ist kein Selbstbedienungsladen

Je mehr Menschen Self-Service Analytics nutzen können, desto mehr werden diese Gelegenheit auch wahrnehmen. Das bedeutet jedoch nicht, dass es sich hier um einen Selbstbedienungsladen für Daten handelt. Die Befähigung von immer mehr Benutzer:innen mit Self-Service erfordert verantwortungsvolle Governance, um...

- zu verhindern, dass die Daten außer Kontrolle geraten
- behördliche Anforderungen einzuhalten
- Vertrauen in die Daten aufrechtzuerhalten, die für geschäftliche Entscheidungen verwendet werden

Es ist wichtig, das richtige Gleichgewicht zwischen dem Schutz von Datenbeständen und der Befähigung von Benutzer:innen zu finden, zusammenzuarbeiten und Nutzen aus den Daten zu ziehen.



Unternehmen müssen das richtige Gleichgewicht zwischen dem Schutz von Datenbeständen (aus einer Governance- und Sicherheitsperspektive) und der Befähigung von Benutzer:innen finden, zusammenzuarbeiten und Nutzen aus den Daten zu ziehen. Es gibt drei Möglichkeiten, dieses Gleichgewicht zu erreichen.

Silos reduzieren

Wenn Datenspeicherung und -verarbeitung in der Cloud mit nahezu unbegrenzter Skalierbarkeit zentralisiert werden und wenn Endbenutzer:innen berechtigt sind, ihre eigenen Daten in die Cloud zu bringen, verhindern Sie, dass Datensilos sich vermehren. Benutzer:innen sammeln Datenextrakte, führen ihre eigenen Vorbereitungsroutrinen aus und erstellen ihre Berichte in und aus der Cloud, anstatt Daten zu extrahieren und in Tabellen zu duplizieren.

Datenkataloge verwenden

Gemeinsam genutzte Ressourcen, wie ein zentraler Katalog oder ein Glossar, mit denen Datendefinitionen, Metadaten und Kenntnisse über die Datenherkunft verwaltet werden können, helfen Benutzer:innen dabei, Daten schneller zu finden und Unternehmen in die Lage zu versetzen, Datenquellen zu verwalten und ihren Lebenszyklus zu überwachen. Mit Lösungen für Machine Learning (ML) und künstliche Intelligenz (KI) kann die Erfassung und Verwaltung von Metadaten und damit zusammenhängendem Wissen über die Daten automatisiert werden.

Datenherkunft nachverfolgen und dokumentieren

Datenherkunft – d. h. wie und von wem Daten verwendet und umgewandelt wurden – ist wichtig für gesetzlich vorgeschriebene Berichte und Audits. Sie ist auch wichtig für Entscheider:innen, die die Historie der Daten hinter Analysen, Visualisierungen und vorausschauenden Empfehlungen sowie die Auswirkungen neuer Anforderungen auf die Datenpipeline zur Erstellung der Analyse verstehen müssen.

Lektion 4 Überprüfen Sie Rollen und Verantwortlichkeiten

Self-Service Datenvorbereitung lädt Unternehmen dazu ein, die Rollen und Verantwortlichkeiten der wichtigsten Stakeholder in der Wertschöpfungskette der Datenanalyse zu überprüfen und den starren Datenpipeline-Prozess als flexibleren und agileren Ablauf neu zu gestalten. Jeder Stakeholder nutzt die Self-Service Datenvorbereitung unterschiedlich für seine Anforderungen, während er für die Bereitstellung von Analysen ein gemeinsames Framework für Design, Kommunikation und Zusammenarbeit verwendet.

Die Self-Service Datenvorbereitung bietet Data Engineers, Data Analysts, Data Scientists, Data Architects sowie analyseorientierten Führungskräften und Managern neue Möglichkeiten zur Zusammenarbeit.

DATA ANALYSTS stehen, neben Business Analysts, Projektmanagern und ähnlichen Positionen, in der Regel in engem Kontakt mit Geschäftsanwendenden. Sie untersuchen häufig Rohdaten, um herauszufinden, was für die Beantwortung geschäftlicher Fragen nützlich sein kann. Ihr Ziel ist es, so schnell und einfach wie möglich Antworten zu erhalten.

Cloud Dataprep by Trifacta eignet sich ideal für Data Analysts, da es einfach zu implementieren ist und keine große technische Erfahrung in der Programmierung erfordert. Data Analysts können frühzeitig in den Datenvorbereitungsprozess einbezogen werden, indem sie Daten untersuchen und Prototypen erstellen, die den Geschäftsanforderungen entsprechen. Wenn Automatisierung für Nachhaltigkeit und Wiederholbarkeit erforderlich ist, können sie mit Data Engineers zusammenarbeiten, um die End-to-End-Datenpipeline zu orchestrieren.



DATA ENGINEERS entwickeln, erstellen und verwalten Datenverarbeitungsprozesse und Datenarchitektur, um Data Analytics und Data Science zu unterstützen. Sie sind eng an der Umwandlung von Daten beteiligt, einschließlich der Untersuchung und Profilierung von Rohdaten. Die wichtigsten Ziele von Data Engineers sind die Optimierung und Automatisierung von datenbezogenen Prozessen, damit sie mehr davon verwalten können.

Cloud Dataprep by Trifacta eignet sich ideal für Data Engineers, da sie damit die verschiedenen Datenströme, die sie oder Datenanalysten entwerfen, operationalisieren und überwachen können. Es erleichtert Dateningenieuren die Zusammenarbeit mit allen Stakeholdern, um die Anforderungen an die Dateninfrastruktur zu verstehen und Benutzenden Anleitung zu geben, wie sie Daten besser erkunden, analysieren, modellieren und nutzen können.

DATA SCIENTISTS wenden Fachkenntnisse an, um Algorithmen zu entwerfen und zu modellieren, indem sie Machine Learning und künstliche Intelligenz nutzen. Bis zu 80 Prozent ihrer Zeit verwenden sie jedoch auf Routineaufgaben zur Datenvorbereitung, sodass wenig Zeit für Innovationen bleibt.

Cloud Dataprep by Trifacta eignet sich ideal für Data Scientists, da es sich um eine Cloud-native Datenvorbereitungslösung handelt. Sie vereinfacht routinemäßige Datenvorbereitungsaufgaben, sodass diese an schneller verfügbare und kostengünstigere Ressourcen delegiert werden können.

Lektion 5

Jeder trägt Verantwortung für die Datenqualität

Alle Stakeholder müssen wissen, dass die Daten, die ihre Analysen, Visualisierungen und Algorithmen vorantreiben, verlässlich sind. Das Ergebnis eines Prognosemodells oder einer analytischen Erkenntnis ist nur so gut wie die Daten, mit denen es gespeist wird.

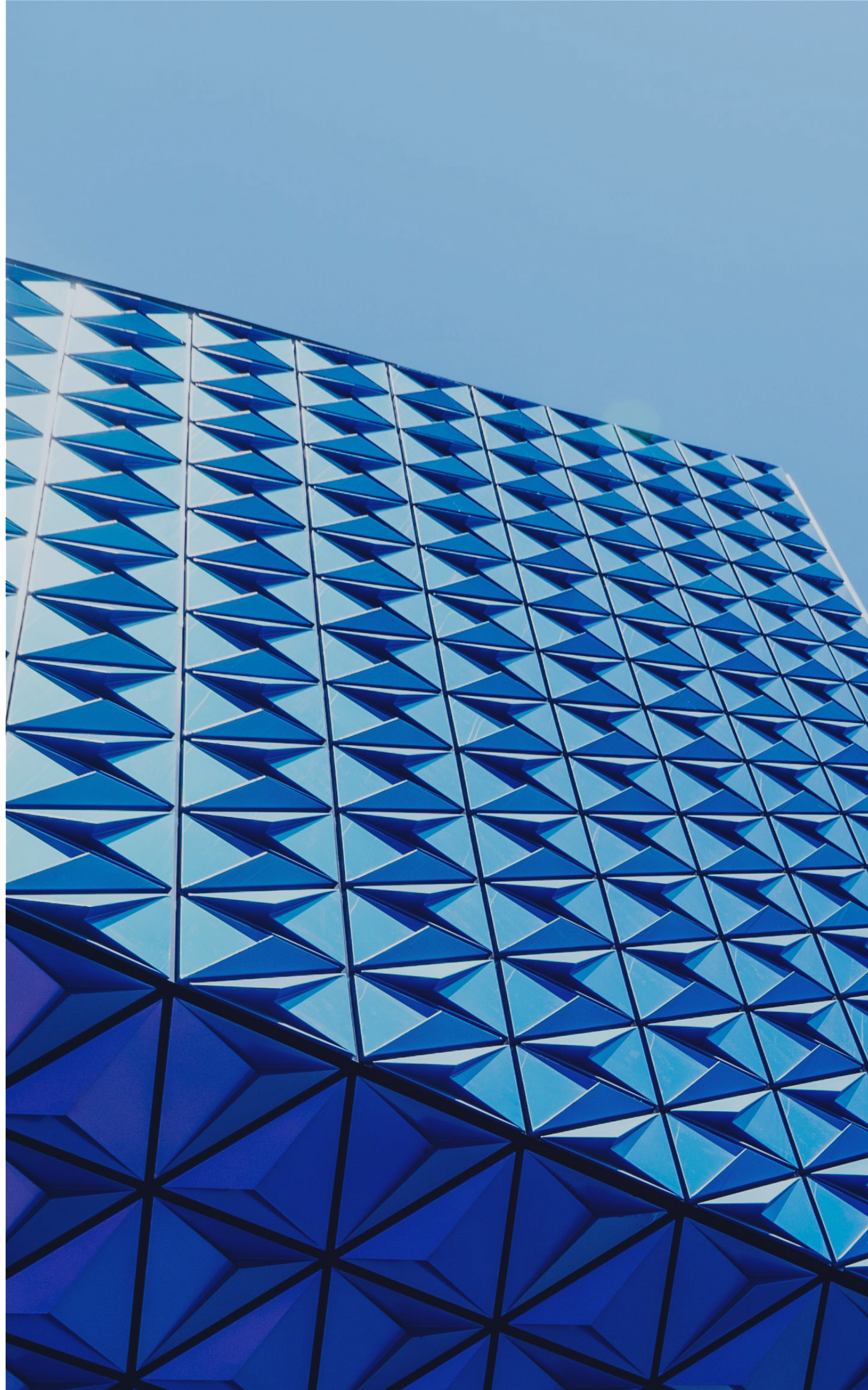
Jedes Unternehmen sollte es zu einer Priorität machen, dass Daten sauber und verlässlich sind.

Es wäre zwar ideal, wenn die Datenqualität auf Unternehmensebene angegangen werden könnte, aber es ist wenig realistisch. Das Datenvolumen und die enorme Vielfalt an Datenquellen (sowohl intern als auch extern), Typen und Kontext erschweren es, die Datenqualität zu bewerten, Fehler zu korrigieren und neue Daten auf Qualitätsprobleme zu prüfen. Neue Data Repositories, wie Cloud Data Lakes und moderne Data Warehouses, stellen komplexere Herausforderungen bei der Datenqualität dar.

Die Verbesserung der Datenqualität in der Datenvorbereitungsphase anzugehen, ist effektiver. Cloud Dataprep by Trifacta verbessert die Genauigkeit, Konsistenz und Vollständigkeit von Daten, indem ML und KI zur Automatisierung von Datenbereinigungsverfahren angewendet werden. Dank Automatisierung können sehr große Data Repositories verarbeitet und Datenwerte, die falsch, ungültig, fehlend oder nicht übereinstimmend erscheinen, schnell erkannt werden. Abweichende Elemente, die eine engere Prüfung rechtfertigen, werden automatisch gekennzeichnet.

Automatisierte Datenprofilierungs- und -bereinigungsroutinen erkennen Inkonsistenzen zwischen Quellen, die in einen Cloud-Datenspeicher integriert werden, heben mögliche Datenduplikate hervor und empfehlen, wie Datenqualitätsprobleme visuell durch codefreie, automatisierte Umwandlungen korrigiert werden können.

Wenn neue strukturierte, unstrukturierte oder halbstrukturierte Daten aufgenommen und in die Cloud integriert werden, wird die Datenqualität kontinuierlich überprüft. Kontinuierliche Überprüfung bedeutet, dass Benutzer:innen nicht bis zum Ende eines Überprüfungsprozesses warten müssen, um Ergebnisse anzuzeigen und zu testen – eine Verzögerung, die sich mit modernen agilen Entwicklungsmethoden nicht vereinbaren lässt.



Fazit

Alteryx hat vielen Unternehmen dabei geholfen, mit Self-Service Analytics in der Cloud erfolgreich zu sein. Aus dieser Erfahrung haben wir fünf wichtige Lektionen gelernt:

1. Die Verwendung eines einfachen Lift-and-Shift-Ansatzes, mit dem lokale Analysen in die Cloud verschoben werden, funktioniert nicht.

Entscheiden Sie sich für eine Cloud-Plattform, die speziell für die Cloud entwickelte Analysekomponenten bietet, wie die Google Cloud Platform mit ihrer umfassenden nativen serverlosen Smart Analytics Suite.

2. ETL-Technologien können Reibungsverluste durch die Verschiebung von lokalen Daten in Cloud-Repositorys minimieren, aber ETL ist zu komplex und unflexibel für Self-Service Analytics. Eine bessere Alternative ist die Einführung von Self-Service-Lösungen zur Datenvorbereitung, wie z. B. Cloud Dataprep by Trifacta, um allen geschäftlichen und technischen Anwendenden die Möglichkeit zu geben, die Daten zu nutzen, die sie für Analysen benötigen.

3. Self-Service ist kein Selbstbedienungsladen ohne Regeln. Damit Self-Service Analytics erfolgreich umgesetzt werden kann, müssen Unternehmen das richtige Gleichgewicht zwischen sorgfältiger Governance und Offenheit finden.

Eine gute Kenntnis der Datenherkunft und die Verwendung eines Geschäftsdatenglossars sind für den Erfolg von entscheidender Bedeutung.

4. Führen Sie neue agile Prozesse ein, die die Zusammenarbeit und eine gemeinsame Sprache unter allen Stakeholdern fördern, um Daten in ein wertvolles Asset für ihre Analysen zu wandeln.

5. Jede Person trägt Verantwortung für die Datenqualität. Cloud Dataprep by Trifacta bietet eine automatische Bewertung der Datenqualität sowie kontinuierliche Überprüfung und Lösung, um das Vertrauen in die Daten zu stärken, die für Self-Service Analytics verwendet werden.

alteryx

ÜBER ALTERYX

Alteryx, das Unternehmen für Analytics Automation, hat es sich zum Ziel gesetzt, jedem Menschen die Möglichkeit zu geben, Daten in einen Durchbruch zu verwandeln. Alteryx vereint Analytics, Data Science und die Automatisierung von Geschäftsprozessen in einer End-to-End-Plattform, um die digitale Transformation zu beschleunigen und die Zukunft von Analytics Automation mitzugestalten. Weltweit nutzen Unternehmen jeder Größe Alteryx, um hochwirksame Geschäftsergebnisse zu erzielen und ihre moderne Belegschaft schnell weiterzubilden. Für weitere Informationen besuchen Sie www.alteryx.com/de.

Alteryx ist eine eingetragene Marke von Alteryx, Inc. Alle anderen Produkt- und Markennamen sind möglicherweise Markenzeichen oder eingetragene Markenzeichen ihrer jeweiligen Eigentümer:innen.

Alteryx Analytics Cloud

Mit Alteryx als Teil der Alteryx Analytics Cloud können Sie Ihre Daten schnell und effizient mit automatisierten Data-Pipelines verknüpfen, profilieren, umwandeln und bereitstellen. No-Code oder Low-Code – wir haben die richtige Lösung für Sie. Prüfen Sie die Datenqualität visuell, beschleunigen Sie die Datenvorbereitung für Analytics, und erstellen und implementieren Sie Data-Pipelines in Sekundenschnelle.

[MEHR ERFAHREN >](#)

Alteryx Machine Learning Plattform

Skalieren Sie Data Science mit automatisiertem Machine Learning (AutoML) und Feature Engineering unternehmensweit, damit Fachkräfte und Data Scientists gleichermaßen in der Lage sind, schneller Erkenntnisse zu gewinnen.

[MEHR ERFAHREN >](#)

Alteryx Designer

Automatisieren Sie jeden Schritt in der Analyse, einschließlich Datenvorbereitung, -zusammenführung, Berichterstellung, Predictive Analytics und Data Science. Greifen Sie auf beliebige Datenquellen, Dateien, Anwendungen oder Datentypen zu und erleben Sie die Einfachheit und Leistungsfähigkeit einer Self-Service-Plattform mit mehr als 300 Automatisierungsmodulen. Beginnen Sie jetzt mit der Erstellung interaktiver Ergebnisse.

[MEHR ERFAHREN >](#)

Alteryx Intelligence Suite

Egal ob Einsteiger:in oder Profi: Erstellen Sie in Minutenschnelle innovative Machine Learning-Modelle. Die Alteryx Intelligence Suite wurde entwickelt, Ihnen dabei zu helfen, in Ihren Daten verborgene Erkenntnisse zu gewinnen. Von der Analyse strukturierter Daten bis hin zum Erschließen von in Texten und Dokumenten verborgenen Erkenntnissen: Mit der Intelligence Suite können Sie selbst die schwierigsten Probleme Ihres Unternehmens lösen.

[MEHR ERFAHREN >](#)

Alteryx Server

Skalieren, teilen und steuern Sie Ihre analytischen Workflow-Prozesse, Modelle und Daten mit Alteryx Server. Erstellen Sie Analyse-Workflows mit Alteryx Designer und veröffentlichen Sie diese auf Alteryx Server, um automatisierte Berichte und Ergebnisse zu planen. Dank integrierter Data Governance, zentral verwalteter Sicherheit und hoher Verfügbarkeit können Unternehmen Analysen abteilungsübergreifend und unternehmensweit skalieren.

[MEHR ERFAHREN >](#)