



Drei grundlegende Veränderungen in modernen Datenumgebungen

und ihre Bedeutung für
IT-Führungskräfte



Inhalt

Das Problem hat sich nicht geändert, es ist nur komplizierter geworden	3
1. Pipelines statt Buckets	5
2. Bedarfsorientierte Datenlandezonen	7
3. Vom Datenwächter zum Datenmentor	10
Das Problem ist komplizierter geworden, aber nicht unüberwindbar	12



Das Problem hat sich nicht geändert, es ist nur komplizierter geworden

Zuverlässige Daten für fundierte Entscheidungen in Unternehmen bereitstellen – diese Aufgabe hat sich seit Jahrzehnten nicht grundlegend geändert. Trotz der massiven technologischen Fortschritte und neuer taktischer Herangehensweisen verfolgen IT-Organisationen heute weiterhin dasselbe Ziel: Daten ab dem Moment ihrer Erstellung weiterzuleiten und den Entscheidungsträgern verfügbar sowie verständlich zu machen, wenn sie diese benötigen.

Zwar ist das Ziel gleich geblieben, doch heute haben sich die Hürden für das erfolgreiche Erstellen und Verwalten einer einzigen zuverlässigen Datenquelle für Analysen im Unternehmen exponentiell vervielfacht.

Das größte Hindernis in modernen Datenumgebungen stellen dabei in den letzten Jahren neue Datenquellen dar, die nie dagewesene Datenmengen mit oftmals geringfügiger (oder gar keiner) Strukturierung enthalten. Von Click-streams, Serverprotokollen und Quellen aus sozialen Medien bis hin zu Geräte- und Sensordaten – der Ansturm von Daten aus solchen Kanälen kann buchstäblich überwältigend sein. Aus wirtschaftlicher und leistungsbezogener Sicht sind herkömmliche Enterprise Data Warehouses (EDWs) einfach nicht in der Lage, diese Welle an Daten zu bewältigen.

Dies hat zu einem kompletten Umdenken in Bezug auf die Datenerfassung und Analysestrategien geführt und zu einer neuen Generation von Datenspeicherlösungen, bei denen schemalose Erfassung, Skalierbarkeit der Hardware und Verlagerung der Rechenleistung näher an die Datenspeicher selbst (wenn nicht sogar zu deren Integration) im Vordergrund stehen.

Im Vergleich mit relationalen Datenbanken sind diese neueren, nicht relationalen Lösungen noch relativ jung. Sie haben in den letzten Jahren jedoch deutlich an Boden gewonnen und sich rasch weiterentwickelt, um mittlerweile einige der größten und komplexesten Unternehmen der Welt zu unterstützen. Auch wenn damit vorrangig eine Ergänzung der vorhandenen EDW-Infrastrukturen beabsichtigt war, müssen die IT-Abteilungen so ein viel komplexeres Datenumfeld verwalten.

Eine weitere Hürde im Rahmen der Bemühungen von IT-Abteilungen, eine intakte Datenumgebung zu gewährleisten, ist die Verfügbarkeit von Daten aus Cloud-Anwendungen. Viele Unternehmen nutzen Anwendungen wie Google Analytics, Salesforce, Netsuite, Zendesk und andere als Kernbestandteil ihrer Infrastruktur.



Die generierten Daten haben eine kritische Bedeutung für das Berichtswesen im Unternehmen. Eine Grundanforderung für die IT ist, Daten aus diesen Cloud-Lösungen einzubinden und für das Unternehmen zugänglich zu machen.

Herkömmliche EDWs dienen nicht mehr als einziges Datenziel und so wird die Frage „wann, wo, wie und ob“ Daten aus Cloud-Anwendungen in die Datenumgebung des Unternehmens übernommen werden sollen, beständig und heftig diskutiert.

Nachdem schließlich Selfservice-Analytics für Unternehmen jeder Größenordnung zum Standard geworden sind, führen immer mehr technisch nicht qualifizierte Anwender (also solche ohne offizielle IT-/Datenschulung) Data-Discovery- und Berichtsaufgaben durch. Dazu gehören manchmal sogar Aufgaben der Vorbereitung und der erweiterten Datenanalyse. Bei Unternehmen, die sich diesem Trend anschließen, zeigt sich oftmals eine drastische Reduzierung (oder sogar Eliminierung) der Zuständigkeiten der IT für Analytics. Diese Verlagerung ist ein wichtiger Baustein für den Gesamterfolg von Unternehmen mit datengesteuerter Ausrichtung, erhöht aber den Druck auf die IT-Organisation, einen breiteren Datenzugriff möglich zu machen. Dabei muss gleichzeitig gewährleistet sein, dass die Technologie die geschäftlichen Anforderungen als auch die der IT nach Sicherheit und Governance erfüllt.

Um die neuen Herausforderungen zu meistern, führen viele IT-Organisationen neue Technologien und Strategien sehr schnell ein. Sie erkennen allerdings nicht, wie diese Hürden die erforderliche Herangehensweise der IT-Abteilungen zum Verwalten von Daten „von der Erstellen bis zu Nutzung“ tatsächlich umgekrempelt haben. Big Data-Lösungen, die Integration von Cloud-Daten und Selfservice-Analytics sind alles Antworten auf größere Technologieprobleme, um sie allerdings in einem Unternehmen effektiv bereitzustellen, muss sich die IT-Strategie ändern.

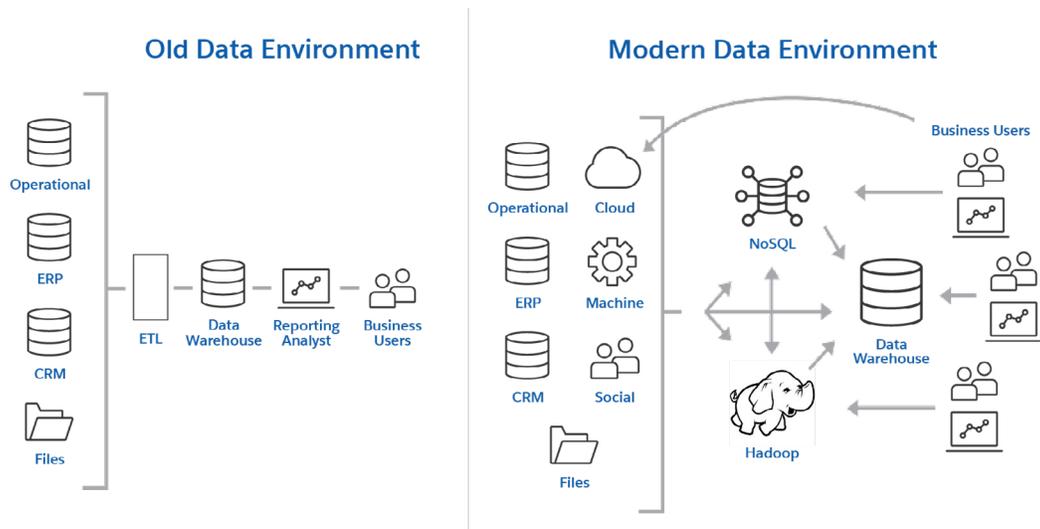
In diesem Whitepaper werden drei grundlegende Veränderungen in der Sichtweise auf die moderne Datenumgebung vorgestellt, die IT-Führungskräfte verstehen müssen, damit sie die datengesteuerte Entscheidungsfindung in ihrer Organisation unterstützen können.



1. Pipelines statt Buckets

Das Enterprise Data Warehouse ist kein Auslaufmodell. Es sind einfach nur ein paar neue Freunde dazugekommen. Von Hadoop-Clustern bis hin zu NoSQL-Datenbanken ist das relationale EDW nicht mehr der einzige offizielle Speicherort für Daten.

Der Prozess der Datenverlagerung ist somit auch nicht mehr unbedingt an einen zentralen Standort gebunden. Tatsächlich sind moderne Datenplattformen heute mehr mit neuronalem Gewebe vergleichbar, das die verschiedenen Bereiche des „Datengehirns“ eines Unternehmens verbindet. Zusammen mit den inzwischen allgegenwärtigen Cloud-Lösungen, die es für Infrastrukturen und Dienste möglich machen, Pipeline-/ETL-Projekte innerhalb von Stunden einzurichten, ergibt dies einen Ansatz, mit dem Daten wie nie zuvor in einem Unternehmen verteilt werden können.



Leider verpassen viele IT-Organisationen diese Gelegenheit, da in vielen Unternehmen auf überholte Konzepte der Datenverwaltung zurückgegriffen wird, die noch auf einzelne „Daten-Buckets“ setzen.

Der Sprung von einer „Bucket-zentrierten“ Denkweise zum „Pipelineansatz“ gelingt, wenn man akzeptiert, dass nicht alle datenbezogenen Fragen in einem Unternehmen aus einer Datenquelle beantwortet werden können. Ein EDW besitzt möglicherweise nicht ausreichend Kapazität für detaillierte Daten, um z. B. E-Mails aus einem Zeitraum von fünf Jahren aufzuschlüsseln, die direkt aus den Protokollen des Nachrichtenservers entnommen werden. Und eine Hadoop-Bereitstellung kann vielleicht innerhalb von Sekundenbruchteilen keine Antworten auf abgefragte Umsatzdaten liefern, die für eine Analyse in Echtzeit benötigt werden.

In der Praxis sind für geschäftliche Probleme sowohl der Detailgrad der Daten als auch die Abfragegeschwindigkeit aus einer oder mehreren Quellen wichtig – allerdings zu unterschiedlichen Zeitpunkten und in verschiedenen Abfolgen, um ein Datenprojekt abschließen zu können.

Unter diesen Voraussetzungen lautet die Frage an die IT also nicht mehr „In welchen Bucket gehören diese Daten?“, sondern „Welche Datenstufe ist erforderlich?“ und „Wie lasse ich Anwender bequem zwischen beiden wechseln?“.

Im Endeffekt geht es hierbei um Pipelines, da der eigentliche Speicherort einer Datenquelle sich ändern kann, wenn das Projekt bis zum Abschluss unterschiedliche Stufen durchläuft. Der Pipelineansatz geht außerdem weit über die herkömmliche Datenintegration hinaus und reicht bis zu organisatorischen Workflows.

Wie erhalten Geschäftsanwender Antworten auf ihre Fragen an Daten?

Die meisten IT-Organisationen sind sich der Tatsache bewusst, dass über das ganze Unternehmen verteilt einzelne „inoffizielle“ Daten-Repositorys in Tabellen angelegt werden. Bislang war man aber nicht in der Lage, diese Ausbreitung einer isolierten, keinen Regeln unterworfenen lokalen Datenverwaltung einzudämmen. Nutzer im Unternehmen gewinnen oft den Eindruck, die einzige praktische Lösung bestehe darin, die Datenumgebung des Unternehmens zu umgehen – und genau das tun sie dann auch.

Wenn IT-Organisationen den kompletten Ablauf nachvollziehen können, wie Geschäftsanwender Daten ermitteln, aufbereiten, analysieren und präsentieren, und dabei zum Unterstützer werden (wobei sie gleichzeitig Hilfestellung bei der Governance leisten), dann – und nur dann – gelingt die Umstellung auf einen Pipelineansatz. Wenn sich die Ansicht durchsetzt, dass die Bereitstellung von Daten – sowohl in der herkömmlichen IT-Bedeutung im Sinne von Batches als auch bezogen auf Ad-hoc-Abfragen der Benutzer im Unternehmen – wichtiger ist als ein sicherer Datenbunker, hat sich ein entscheidender und zentraler Wandel für Technologieverantwortliche vollzogen.

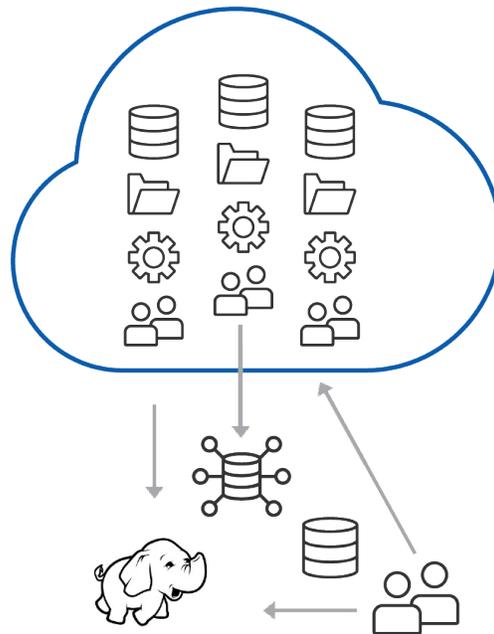


Aus globaler Sicht wird es immer entscheidender, über all diese Informationen zu verfügen und sie verstehen zu können. Bei dem System, das wir jetzt mit unserem logischen Data Warehouse und mit Tableau aufgebaut haben, können die Benutzer zusammenarbeiten und es sind alle diese Informationen verfügbar.

– Tim Nall, CIO,
[Brown-Forman](#)

2. Bedarfsorientierte Datenlandezonen

Daten aus Cloud-Anwendungen müssen nicht immer direkt in das Enterprise Data Warehouse übernommen werden. Entsprechende Entscheidungen sollten stets die Anforderungen des Unternehmens berücksichtigen. Genauer gesagt sollte eine Entscheidung das abbilden, was nötig ist, damit die Daten für das Unternehmen hilfreich sind.



Ein weit verbreiteter Irrtum bei IT-Organisationen, auch bei den Befürwortern des Ansatzes zur Abschaffung einzelner Buckets mit Daten, ist, dass Daten aus Webanwendungen direkt in das EDW eingebunden werden sollten. Dies erscheint auf den ersten Blick sinnvoll, da viele wichtige Cloud Anwendungen für Unternehmen zentraler Bestandteil der Vertriebs-, Marketing- und Supportinfrastruktur sind. Hierbei handelt es sich praxisrelevante, häufig aufgerufene und oftmals gut strukturierte Daten. Der ROI beim Platzieren dieser Datenbestände im EDW scheint vertretbar, obwohl dies zum höchsten Preis pro Byte führt.

IT-Organisationen müssen aber die folgenden kritischen Fragen beantworten:

- Sind die Daten aus der Cloud direkt für die Analyse verfügbar?
Anders gesagt: Sind die Daten aufbereitet und sollten wir uns die Mühe machen, sie zu verlagern?
- Ist der tatsächliche Wert dieser Cloud-Daten erst realisierbar, wenn sie mit weiteren Daten kombiniert werden?

Abhängig von den Antworten auf diese Fragen kann der von der IT-Organisation festgelegte Speicherort für Cloud-Datenbestände (sofern diese Daten überhaupt an einen internen Speicherort verschoben werden) von Unternehmen zu Unternehmen und von Anwendung zu Anwendung unterschiedlich sein.

Jede IT-Organisation, die für eine Salesforce.com-Bereitstellung einer gewissen Größe verantwortlich ist, wird umgehend bestätigen, dass das Speichern nicht deduplizierter Datensätze mit Verkaufschancen im EDW bedeutet, einen unheilvollen Weg einzuschlagen. Von fehlerhaften Provisionszahlungen bis zu deutlich überschätzten Umsatzprognosen sind diese Probleme nur die Spitze des Eisbergs, wenn es um die Verarbeitung nicht ordnungsgemäß aufbereiteter Vertriebsdaten geht.



Indem wir die Daten unseren Endbenutzer zur Verfügung gestellt haben, haben wir ihnen die Möglichkeit gegeben, Basisberichte selbstständig zu erstellen. Diese Mitarbeiter waren am nächsten an den Produkten und den Anwendungsänderungen dran und konnten schneller erkennen, wo Anpassungen erforderlich waren.

– Sharon Graves, Enterprise Data Evangelist, [GoDaddy](#)

Genauso können Website- und Marketingdaten aus Cloud-Quellen wie Google Analytics, Eloqua und Marketo negative Auswirkungen auf die Fähigkeit eines Unternehmens haben, z. B. die Lead-Entwicklung nachzuverfolgen oder die Akquisitionskosten für Kunden zu berechnen, wenn kein Verlass auf die Genauigkeit der Datensätze ist.

Dieses Genauigkeitsproblem nimmt exponentiell zu, wenn Quellen zusammengeführt werden müssen. Das ist häufig der Fall, wenn ein Unternehmen eine umfassende Darstellung des Kundentrichters benötigt – vom ersten Klick des Kunden auf der Website bis zur Kaufentscheidung. Auch wenn fast alle IT-Organisationen die Bedeutung bereinigter Daten, aus der Cloud oder anderen Quellen, verstehen, betrachten sie die Daten nicht im Zusammenhang mit ihrem ersten Landepunkt nach dem Auslesen aus einer Webanwendung.



IT-Organisationen mit einer Datenumgebung, die auf mehreren Buckets und Pipelines basiert, bewerten die Einbindung von Cloud-Daten auf der Grundlage des Werts, den die Datenbestände für das Unternehmen zu dem Zeitpunkt haben, an dem sie in das Unternehmen lokal übernommen werden. Bei weitgehend aufbereiteten Cloud-Datenquellen, die nach ihrer Übernahme einen extrem hohen Wert für das Unternehmen darstellen, sind IT-Organisationen gut beraten, sie dort zu speichern, wo am schnellsten darauf zugegriffen werden kann (relationales EDW).

Handelt es sich jedoch um Daten aus Webanwendungen, für die eine umfangreiche und/oder komplexe Verarbeitung anfällt, damit sie für das Unternehmen vertrauenswürdig sind, können IT-Organisationen Umgebungen mit hoher Rechenleistung und geringen Kosten pro Byte einsetzen, z. B. eine Hadoop-Umgebung. Der zweite Ansatz ermöglicht eine maximale Nutzung der Ressourcen für die Aufbereitung und Umwandlung der Daten, ohne die Geschwindigkeit des EDW zu beeinträchtigen. Danach kann die IT gemeinsam mit dem Geschäftsbetrieb entscheiden, ob die aufbereiteten Daten in ein EDW verlagert werden und/oder ob innerhalb der erweiterten Umgebung direkt darauf zugegriffen wird.

Möglicherweise ist es gar nicht erforderlich, Daten aus Cloud-Anwendungen zu verlagern. Immer häufiger stellen Webanwendungen einen bequemen Zugriffspunkt für die Backend-Repositorys bereit, sodass geschäftliche Anwender Selfservice-Tools für Berichte und Analytics verwenden und so eigene Untersuchungen von Live-Daten durchführen können.

Für IT-Organisationen, die zusätzlich eine Zwischenebene für Autorisierung und Governance in einem solchen Szenario einfügen möchten, bieten einige der unternehmensfreundlichen Selfservice-Analytics-Lösungen auch Proxy-Verbindungen. So können alle Anforderungen der IT von einfachem Benutzerzugriff bis hin zu einer hochkomplexen Unternehmenslogik erfüllt werden. Vor diesem Hintergrund lautet die Frage zur Einbindung von Daten aus Cloud-Anwendungen nicht „Wo?“, sondern „Überhaupt erforderlich?“. Gibt es einen zwingenden Grund dafür, die Daten überhaupt zu verlagern (Bereinigung, Mehrwert usw.) oder können sie an ihrem Speicherort verbleiben und potenziell auf der Ebene der Geschäftsanwender eingebunden werden?



Für Tableau haben wir uns entschieden, weil es ein solides Unternehmen ist, das einen klaren Fahrplan für die Zukunft anbietet. Dieser Fahrplan passt zu dem, was ich erreichen möchte. Mit ihm kann ich mich in dem Tempo bewegen, das für mich richtig und sinnvoll ist.

– Steven John, CIO,
[Ameripride](#)



3. Vom Datenwächter zum Datenmentor

Die Entwicklung hin zu Selfservice-Analytics in Unternehmen sollte als Angebot zur Zusammenarbeit mit der IT verstanden werden, nicht als eine Art feindliche Übernahme. Und die IT-Organisationen, die dieses Angebot als erste annehmen, können dabei nur gewinnen.

In gleicher Weise, wie die moderne Datenumgebung erweitert wurde und jetzt mehr als nur relationale Datenspeicher umfasst, besteht auch die Analytics-Umgebung inzwischen aus Tools, die für den normalen Geschäftsanwender enormen Wert bieten. Dies hat zu einer radikalen Abkehr von der herkömmlichen Vorgehensweise geführt, bei der Analytics in den Händen einiger weniger Spezialisten lagen.

A Partnership That Works

IT Role

Security
Data Architecture
Scalability
Training
Corner of Operations
Enablement Intranet

Business Role

Creative Analytic Work
Dining Data Acquisition
Sharing Expertise
Seeking Help When Needed
Evangelism
Catalyzing Action

Enablement → Execution



Infolgedessen werden zukunftsorientierte IT-Organisationen zu Vordenkern in Sachen Daten und Analytics ihrer Unternehmen und sie definieren die Zusammenarbeit mit Geschäftsteams bei der Lösung von datenbezogenen Problemen komplett neu. Diese Wandlung der IT vom „Datenwächter“ zum „Datenmentor“ stellt vielleicht die bedeutsamste Änderung in der modernen Datenumgebung dar. Sie ist ein Schlüsselfaktor für Erfolg oder Misserfolg auf dem Weg zu einem wirklich datengesteuerten Unternehmen.

Mit der neu definierten Beziehung zwischen IT und Geschäftsbetrieb ändert sich auch die generelle Haltung in Bezug auf die Daten-Governance von Ablehnung in Bewilligung. Bislang benötigten Geschäftsanwender eine ausdrückliche Genehmigung, damit der Zugriff auf Daten nicht automatisch verweigert wurde. Das neue Modell geht davon aus, dass jeder Geschäftsanwender im Unternehmen auf alle Daten zugreifen kann, sofern dies keine Verletzung der Compliance-Vorschriften darstellt.

Diese neue Sichtweise bewirkt enorme Veränderungen. Die Auswirkungen zeigen sich darin, wie Geschäftsanwender nach neuen Daten suchen, um neue Probleme zu lösen. Dabei steht die Bereitschaft zur Problemlösung mit neuen Daten in direktem Zusammenhang mit dem vom Benutzer erwarteten Aufwand für den Zugriff auf neue Daten. Ausgestattet mit einer benutzerfreundlichen Selfservice-Analytics-Lösung und bestärkt durch das Wissen, dass die IT-Organisation einen breit angelegten Datenzugriff unterstützt, können die Mitarbeiter im Unternehmen zu Change Agents werden. Als solche machen sie auch deutlich, dass der Kurs hin zu einem datengesteuerten Unternehmen erfolgreich eingeschlagen wurde.

Durch den einfacheren Zugang zu Daten und die Bereitstellung neuer Software ist jedoch das größere Problem nicht gelöst, das lautet, den Geschäftsbetrieb dabei zu unterstützen, verbesserte Entscheidungen durch Daten zu treffen. Jene IT-Organisationen, die erfolgreich daran mitgewirkt haben, in ihren Unternehmen eine Kultur datengesteuerter Entscheidungsfindung zu etablieren, haben dies in ihrer Rolle als Datenmentor möglich gemacht. Oftmals werden Mitglieder des IT-Teams, die zuvor den Analytics-Prozess gesteuert haben, zu Lotsen für die anderen Abteilungen im Unternehmen. Sie besitzen das Know-how, um anderen zentrale Funktionen zu zeigen, z. B. für eine ordnungsgemäße Datenanalyse oder ein präzises Reporting.

Wenn Sie Geschäftsanwendern auf diese Weise Wissen vermitteln, steigert sich die Analyse- und Berichtscompetenz des gesamten Unternehmens um ein Vielfaches. Zudem erzielen Unternehmen damit einen höheren und schnelleren ROI in Bezug auf Selfservice-Analytics-Lösungen. Der wichtigste Punkt ist jedoch, dass die Nutzung von Daten für eine ganzheitliche Betrachtung des Unternehmens allgemeiner Bestandteil der Kultur wird.

Das Problem ist komplizierter geworden – aber nicht unüberwindbar

Die Zuständigkeit von IT-Organisationen für die Steuerung des Datenflusses – von der Entwicklung neuer Lösungen für das Speichern und Verarbeiten von Daten bis zur Festlegung, wann, wo und wie Cloud-Daten in das gesamte Umfeld eingespeist werden – wird immer komplexer. Hinzu kommt der Trend zu Selfservice-Analytics, der diesen Wandel insgesamt verstärkt. Damit vollzieht sich eine tiefgreifende Umwälzung in der Art und Weise, wie Unternehmen mit Daten verfahren.

Die Firmen benötigen die IT als Lotse durch diese Veränderung. Die Unsicherheit der IT-Organisationen ist größtenteils auf die Erkenntnis zurückzuführen, dass es nicht mehr nur darum geht, neue Technologien in vorhandene Geschäftsprozesse einzubinden. Big Data, Cloud-Technologien und die Bereitstellung eines Self-service-Analytics-Modells stellen grundlegende Änderungen der Art und Weise dar, wie IT-Abteilungen moderne Datenumgebungen gestalten müssen.

Für Unternehmen, die bereits damit begonnen haben, ihre Einstellung gegenüber Daten und Analytics zu überdenken, um auf diese Veränderungen zu reagieren, sind die damit verbundenen Hürden Realität – ebenso wie die Erkenntnisse, die sie damit gewinnen. Veränderungen sind schwierig, sowohl für Geschäftsanwender als auch für die IT. Bei der Neudefinition der Beziehung zwischen beiden wird es in jeder Fraktion Widerstände geben. Andere werden Veränderungsbereitschaft an den Tag legen.

Im Zuge der Lockerung starrer Begrenzungen und Vorgehensweisen beim unternehmensweiten Datenzugriff werden IT-Organisationen unerwartet auf Analytics-Befürworter in allen Unternehmensbereichen stoßen, denen daran gelegen ist, zu Vorreitern in einem neuen datengesteuerten Unternehmen zu werden. Umsichtige IT-Führungskräfte stellen diesen neuen Befürwortern ihre erfahrenen Datenexperten als Mentor an die Seite. Die Befürworter werden sich dabei in das ständig weiterentwickelnde Datenumfeld einarbeiten und die IT-Spezialisten werden die geschäftlichen Problemstellungen kennenlernen, die es zu lösen gilt. Diese ersten Schritte sind die Voraussetzung für die Schaffung einer durch Analytics und Data Discovery geprägten Kultur in einem Unternehmen, in dem Entscheidungen auf der Sicherheit und Präzision einer hochmodernen Datenumgebung basieren.



Über Tableau, ein Salesforce-Unternehmen

Tableau hilft Nutzern, ihre Daten sichtbar und verständlich zu machen. Als weltweit führende Analytics-Plattform bietet Tableau Visual Analytics mit leistungsstarken Funktionen für KI, Datenmanagement und Zusammenarbeit. Von Einzelpersonen bis zu Unternehmen jeder Größenordnung nutzen Kunden weltweit begeistert fortgeschrittene Analytics von Tableau für datengesteuerte Entscheidungen, die den Unterschied ausmachen. Weitere Informationen finden Sie unter www.tableau.com/de-de/

START YOUR FREE TRIAL



Ihre Tableau Cloud-Testversion umfasst auch Prep Builder und das Data Management-Add-on

START NOW

Weitere Materialien

[Tableau für die IT](#)

[Tableau Data Management](#)

[Evaluierungsleitfaden für moderne BI und Analytics](#)

[Kontrollierte Selfservice-Analytics in großem Umfang](#)

[Neudefinition der Rolle der IT-Abteilung in einer modernen](#)

[BI-Umgebung](#)

[Tableau für Unternehmen: Ein IT-Überblick](#)





 **tableau**®

