

# Tableau und Big Data: Ein Überblick



# Inhaltsverzeichnis

<b>Big Data heute – Entwicklung, Möglichkeiten, Gefahren</b> .....	<b>3</b>
Datenentwicklung und Analysebedarf .....	3
Big Data: Möglichkeiten und Gefahren .....	4
<b>So arbeitet Tableau mit Big Data</b> .....	<b>5</b>
Das große Ganze .....	5
Datenzugriff und Verbindungen .....	5
Schnelle Interaktion mit allen verfügbaren Daten .....	6
<b>Tableau und das Ökosystem für Big-Data-Analysen</b> .....	<b>7</b>
Cloud-Infrastruktur .....	8
Aufnahme und Vorbereitung .....	8
Speicherung und Verarbeitung .....	9
Abfragebeschleunigung .....	10
Datenkatalog .....	10
<b>Big-Data-Analysearchitekturen</b> .....	<b>10</b>
Beispiele für die größten Cloud-Anbieter .....	11
Beispiele für Tableau-Kunden .....	12
Gängige Muster .....	13
<b>Über Tableau und weitere Materialien</b> .....	<b>14</b>



# Big Data heute – Entwicklung, Möglichkeiten, Gefahren

## Datenentwicklung und Analysebedarf

Daten sind überall – wie auch der Bedarf, auf diese zugreifen zu können und sie zu analysieren. „Big Data“ hat sich wohl als Modewort etabliert, aber die drei Big-Data-Dimensionen – Datenvolumen, Bandbreite und Geschwindigkeit – beziehen sich mehr als jemals zuvor auf Anwendungsfälle der Big-Data-Analyse. Obwohl diese und andere in der Branche diskutierten Dimensionen subjektiv sind (wie Variabilität, Gültigkeit, Wahrhaftigkeit usw.), erinnern sie uns doch daran, dass es sich heutzutage bei Big Data noch immer schlicht um Daten handelt – diese sind aber so komplex geworden, dass Organisationen Innovationen brauchen, um sie effektiv sammeln, kuratieren, verstehen und nutzen zu können.

Der digitale Wandel wirkt sich auf jede Branche und auf Organisationen jeder Größe aus, wobei enorme Datenmengen verschiedenen Ursprungs in vielen Formaten entstehen. Heutzutage sammeln, verarbeiten und analysieren Organisationen vielfältigere Daten als jemals zuvor. Vom schemafreien JSON über verschachtelte Typen in anderen Datenbanken, wie etwa relational und NoSQL, bis zu strukturierten Daten – wie Avro, Parquet, XML, usw. – vervielfältigen sich die Datenformate und die Nutzung der Konnektoren wird entscheidend.

## Organisationen haben häufig eine Kombination aus folgenden Datentypen:

- **Strukturierte Daten** mit vorberechneten Mengen für spezifische Fragen, möglicherweise für die speicherresidente Verrechnung als Datenextrakt aufgerufen und für die Analyse gesammelt. Dies sind normalerweise die präzisesten und am leichtesten zugänglichen Daten, über die eine Organisation verfügt.
- **Halbstrukturierte Daten** (oder Objektspeicher), möglicherweise in relationalen Datenbanken, Data Warehouses oder Datenmärkten. Dabei handelt es sich häufig um regelmäßig erneuerte Geschäftskonzepte für die Entitätenanalyse – bekannte Fragen mit unbekanntem Antworten –, zum Beispiel Transaktionen, Möglichkeiten oder Maßnahmen, die von einzelnen Verkäufern im Zuge der Möglichkeiten getroffen werden.
- **Unstrukturierte Rohdaten** in einem Datensee oder einem Cloud-Speicher. Dies umfasst Stream-Daten, die von Social-Network-Feeds, IoT-Geräten und anderen erstellt wurden. Data Scientists können diese Daten sammeln und umwandeln, aber das volle Potential ist weiterhin unbekannt.

Während für einige Daten noch nützliche Anwendungsfälle gefunden werden müssen, steigt der Bedarf an qualifizierten Angestellten, die diese Daten für die Entscheidungsfindung aufrufen und analysieren. Anwendungen zur Datenanalyse und Visualisierung bewegen sich immer mehr auf die Daten selbst zu. Dies erfordert einen groß angelegten Wechsel zur Cloud, in der die Datenanalyse gemeinsam mit robusten Speicher- und Datenverarbeitungsdiensten stattfinden kann, die eine größere Flexibilität und einen größeren Maßstab ermöglichen. Ob ein Unternehmen nun umfangreich und cloudbasiert Big Data



nutzt oder Daten aktuell nur selten analysiert – alle Unternehmen profitieren erheblich, wenn sie ihren Mitarbeitern in den Geschäfts- und IT-Abteilungen die Möglichkeit geben, Muster zu visualisieren und sie auf die enthaltenen Erkenntnisse hin zu analysieren.

Obwohl moderne Analyseverfahren mehr Geschäftsanwendern aller Qualifikationsstufen zusätzliche Fähigkeiten bieten, erweist es sich doch als eine Herausforderung, Möglichkeiten zu finden, um alle diese Daten zu einer nützlichen Ressource für die gesamte Organisation zu machen. Geschäftliche Anforderungen ändern sich so häufig wie die Daten selbst, weshalb eine agile und anpassbare Big-Data-Strategie und -Architektur notwendig sind. Anstatt monolithische Plattformen mit Schwerpunkt auf die Datenverbindung aufzubauen, wären Organisationen gut beraten, ihren Big-Data-Bereich auszubauen und über die sich entwickelnden Analyseanwendungsfälle nachzudenken. Sie könnten andernfalls den Überblick verlieren.

## **Big Data birgt sowohl Möglichkeiten als auch Risiken**

Datenbestände werden zum wesentlichen Unterscheidungsmerkmal zwischen extrem rentablen und schwächelnden Unternehmen. Angesichts des enormen Volumens, rasanten Wachstums und der Vielfalt der Daten stoßen relationale Datenbankverwaltungssysteme an Kapazitäts- und Kostengrenzen. Zusätzlich zu den Einsparungen bei den Hardware-Kosten dank Vorberechnung und geteilter Berechnung versuchen die Kunden zudem, ihre Daten so wenig wie möglich zu verschieben. Die Infrastruktur ermöglicht es ihnen, Daten auf agile Art und Weise zu bewegen, und hilft, die Lücke zwischen unstrukturierten Rohdaten und Daten, die zur Analyse bereit sind, zu schließen.

Organisationen haben zudem Probleme mit der Konnektivität und Leistung. Selbst bei Optionen für Live-Verbindungen oder speicherresidente Analysen können riesige Datenseen schwerfällig bei der Generierung von Extrakten oder beim Mischen mit anderen Daten sein. Der moderne Ansatz zum Analytics-Selfservice verspricht vor allem Agilität, jedoch können massive Verknüpfungen auf die Datensätze das System ersticken.

Hier müssen IT und Unternehmen zusammenarbeiten, mit einem Bottom-up-Ansatz bestehend aus Fachkräften, die Metadaten, Geschäftsregeln und Berichtsmodelle erstellen. Diese Prozesse müssen fortlaufend wiederholt und verbessert werden, um den stets neuen Anforderungen an das Geschäft gerecht zu werden. In Zeiten der digitalen Transformation steht das Geschäft nie still, deshalb sollte Ihr Big-Data-Analyserahmen es auch nicht.



# So arbeitet Tableau mit Big Data

## Das große Ganze

Grundlage für alles, was wir bei Tableau tun, ist unser Ziel, den Benutzern dabei zu helfen, ihre Daten sichtbar und verständlich zu machen. Tableau ist eine moderne Analytics-Plattform für die digitale Wirtschaft, weil wir fest an die Demokratisierung von Daten glauben. Die Personen, die sich mit den Daten auskennen, sollten auch diejenigen sein, die in Bezug auf die Daten Fragen stellen dürfen. Das bedeutet, dass Wissensarbeiter aller Qualifikationsstufen die Möglichkeit haben sollten, auf ihre Daten zuzugreifen, sie zu analysieren und Einblicke zu bekommen, von überall aus.

Da sich viele Kunden mit vielfältigen Big-Data-Technologien beschäftigen, haben wir unsere technischen Investitionen, Partnerschaften innerhalb des Ökosystems und die allgemeine Vision auf die Entwicklung der Datenlandschaft ausgerichtet. Tableau hat eine umfangreiche Big-Data-Investitionsgeschichte, die ihrer Zeit voraus ist. Diese Investitionen umfassen die Datenkonnektivität zu sowohl Hadoop- und NoSQL-Plattformen sowie umfangreiche Data Warehouses vor Ort und in der Cloud.

“ Wir begannen mit einem sehr eng gefassten Geschäftsanwendungsfall, der sich dann schnell immer weiter ausbreitete. Jeder möchte über Big-Data-Analytics sprechen, doch Tableau macht es einfach.

– ASHISH BRAGANZA, DIRECTOR OF GLOBAL BUSINESS INTELLIGENCE, LENOVO

**Erfahren Sie, wie Lenovo die Effizienz des Berichtswesens in 28 Ländern um 95 % gesteigert hat.**

## Datenzugriff und Verbindungen

Um die Analyse von Daten jeder Größe, jeden Speicherortes und in jedem Format zu ermöglichen, unterstützen wir einen breiten Datenzugriff. Tableau unterstützt gegenwärtig mehr als 75 native Datenkonnektoren sowie zahlreiche weitere Quellen, die über unsere Erweiterungsoptionen zugänglich gemacht werden können. Vor dem Hintergrund, dass immer neue Datenquellen auftauchen, die für unsere Kunden wertvoll sind, werden wir die Händlerkonnektoren in Tableau integrieren und zertifizieren, um einen reibungsloseren Datenzugriff zu gewährleisten. Wir sind der Meinung, dass es viele Datenquellen gibt und immer geben wird, die eine Person nutzen möchte – ob Web-Datenverkehr, Einträge in Datenbanken, Protokolldateien usw.

- **SQL-basierte Anschlüsse** — Tableau verwendet SQL als Schnittstelle zu Hadoop, NoSQL-Datenbanken und Spark. Die von Tableau generierten SQL-Anweisungen entsprechen der Norm ANSI SQL-92. Der Einsatz von SQL lohnt sich, denn die Sprache ist äußerst kompakt (ein einziger Ausdruck), quelloffen, standardisiert und es gibt keine Bibliotheksabhängigkeiten. Zudem ist SQL sehr vielfältig und ausdrucksstark. Mittels SQL lassen sich u. a. Join-Vorgänge, Funktionen, Kriterien, Zusammenfassungen, Gruppierungen und verschachtelte Vorgänge darstellen.



- **NoSQL-Schnittstellen** — Wie der Name schon andeutet, können NoSQL-Datenbanken („nicht nur SQL“) Daten enthalten, die zusätzlich zu relationalen auch in unrelationalen Formaten modelliert sind und weitere Speichertypen unterstützen, einschließlich Spalten, Dokumente, Schlüsselwerte und Grafiken. Ebenso können NoSQL-Datenbanken auch SQL-ähnliche Schnittstellen unterstützen.
- **ODBC** — Die von Tableau verwendeten Treiber nutzen den Programmierstandard Open Database Connectivity (ODBC) als Übersetzungsschicht zwischen SQL- und SQL-ähnlichen Datenschnittstellen, die von diesen Big-Data-Plattformen bereitgestellt werden. Durch die Nutzung von ODBC können Sie auf jede Datenquelle zugreifen, die SQL Standard unterstützt und die ODBC-API implementiert. Für Hadoop sind das Schnittstellen wie Hive Query Language (HiveQL), Impala SQL, BigSQL und Spark SQL. Zur Erzielung der bestmöglichen Leistung erfolgt eine Optimierung der generierten SQL. Aggregationen, Filterfunktionen, und andere SQL-Vorgänge werden auf die Big-Data-Plattformen verlagert.
- **Webdaten-Connector** — Mit dem Tableau Webdaten-Connector SDK können Verbindungen zu Daten aufgebaut werden, die sich außerhalb der bestehenden Konnektoren befinden. Nutzer von Selfservice-Analytics können ihre Big-Data-Analysen mit externen Daten erweitern, indem sie über HTTP eine Verbindung zu fast allen zugänglichen Daten herstellen, einschließlich interner Web-Diensten, JSON-Daten und REST-API.

## Schnelle Interaktion mit allen betreffenden Daten

Wir wollen, dass die Nutzer Zugriff auf alle betreffenden Daten haben, um diese in andere Daten zu integrieren und schneller zu Erkenntnissen zu gelangen. Um die Selfservice-Visual-Analytics mit Big Data möglich zu machen, hat Tableau in mehrere bahnbrechende Technologien investiert.

- **Hyper-Daten-Engine** — **Hyper** ist unsere leistungsstarke speicherresidente Daten-Engine-Technologie, die den Kunden dabei hilft, große und komplexe Datensätze schneller zu analysieren. Die urheberrechtlich geschützte dynamische Code-Generierung und modernste Parallelismustechniken ermöglichen Hyper eine bessere Nutzung moderner Hardware. So werden dreimal schneller Extrakte erstellt und eine fünffache Abfragegeschwindigkeit im Vergleich zur vorherigen Tableau-Daten-Engine erreicht. Hyper kann zudem langsamere Datenquellen aufwerten und beschleunigen, indem ein Datenextrakt erstellt und in die speicherresidenten Daten eingefügt wird.
- **Hybride Datenarchitektur** — Tableau kann sich live mit Datenquellen verbinden oder Daten (bzw. eine Teilmenge) speicherresident einbringen. Sie können entsprechend Ihren Bedürfnissen zwischen diesen Modi vor- und zurückschalten. Unser hybrider Ansatz für den Datenzugriff ermöglicht den Anwendern viel Flexibilität und kann bei der Optimierung der Abfrageleistung helfen.

- **VizQL™** — Das Herzstück von Tableau ist eine unternehmenseigene Technologie, die die interaktive Datenvisualisierung zu einem wesentlichen Bestandteil beim Verstehen von Daten macht. Mit herkömmlichen Analysetools müssen Daten in Zeilen und Spalten eingetragen und ein Teildatensatz ausgewählt werden. Die Daten müssen in einer Tabelle zusammengefasst und daraus eine Grafik erstellt werden. Bei VizQL werden diese Schritte übersprungen und es wird sofort eine visuelle Darstellung Ihrer Daten erstellt, damit Sie bereits während der Analyse ein grafisches Feedback erhalten. VizQL ermöglicht Ihnen eine grenzenlose Untersuchung Ihrer Daten, um die beste Darstellungsform zu finden – und dank der Möglichkeit, Maßnahmen unbegrenzt rückgängig zu machen, wird jeder Fehler verziehen. In diesem Zyklus der visuellen Analyse ist Praxislernen angesagt. Nutzer können bei Bedarf mehr Daten hinzufügen und erhalten letztendlich tiefere Einblicke. Somit ist nicht nur das allgemeine Erlebnis besser, sondern auch der Zugang für Nutzer unterschiedlichen Kenntnisstandes deutlich einfacher als beim Codieren von Dashboards.

 Bei Lenovo können Sie sogar in Echtzeit mit dem Datensatz interagieren und sind in der Lage, diesen zu analysieren und anschließend entsprechend Ihren Wünschen in wenigen Minuten vorzustellen.

—JAMIE FAN, PRODUCT ANALYTICS LEAD, GRAB

**So analysiert man bei Grab Millionen von Datenreihen und verbessert das Kundenerlebnis**

## Tableau und das Ökosystem für Big-Data-Analysen

Eine moderne Analytics-Plattform wie Tableau ist möglicherweise der Schlüssel, um das Potential von Big Data anhand wichtiger Einblicke zu entfalten. Trotzdem handelt es sich nur um eine der kritischen Komponenten einer vollständigen Big-Data-Architektur. Eine ganze Big-Data-Analytics-Pipeline zu gestalten, kann eine Herausforderung sein. Die gute Nachricht ist, dass Sie weder das gesamte Ökosystem ausbauen müssen, bevor Sie beginnen, noch müssen Sie jede einzelne Komponente für eine ganze Strategie integrieren, um voranzukommen.

Tableau passt ausgezeichnet in das Big-Data-Paradigma, da Flexibilität priorisiert wird. Das bedeutet, Nutzer haben die Möglichkeit, Daten plattformübergreifend zu bewegen, auf Anfrage die Infrastruktur anzupassen, neue Datentypen zu nutzen und neue Anwender und Anwendungsfälle aufzunehmen. Wir glauben, dass die Nutzung einer Big-Data-Analytics-Lösung nicht Ihre Infrastruktur oder Strategie diktieren sollte, jedoch sollte sie sich positiv auf Ihre bereits vorgenommenen Investitionen auswirken, einschließlich solcher mit Partnertechnologien innerhalb des Big-Data-Ökosystems.



## Cloud-Infrastruktur

Organisationen verlegen Ihre Geschäftsprozesse und Infrastruktur zunehmend in die Cloud. Während Cloudbasierte Infrastrukturen und Datendienste einige der Haupthindernisse bei Hadoop-Datenseen vor Ort beseitigt haben, sind Cloudbasierte Big-Data-Analytics-Lösungen heute leichter als jemals zuvor zu implementieren und zu verwalten.

Hadoop hat mit seiner kraftvollen Kombination aus geringen Kosten, skalierbarem Speicher (Hadoop Distributed File System-HDFS), spezialgefertigten Verarbeitungsfunktionen (erst MapReduce, dann im Laufe der Zeit Hive, Impala und Spark) und einen geteilten Datenkatalog die Grundlagen für moderne Datenseen gelegt.

Heute können einmal zugeordnete Speicher- und Verrechnungsdienste nach Bedarf und unabhängig in der Cloud skaliert werden. Auch Ressourcen können einfacher ausgebaut oder reduziert und auf Anfrage mit einem Preis versehen werden. Allgemein sorgt die Cloud für eine höhere Effizienz, Verwaltung und Koordination aller Dienste.

[Hier erfahren Sie mehr über die vielen Vorteile dieses Produkts](#) von Josh Klahr, VP of Product at AtScale.

Tableau liefert Schlüsselintegrationen mit cloudbasierten Technologien, die bereits von vielen Unternehmen und Organisationen genutzt werden, einschließlich [Amazon Web Services](#), [Google Cloud Platform](#) und [Microsoft Azure](#).

## Aufnahme und Vorbereitung

Bei modernen Aufnehmen-und-Laden-Mustern ist das Ziel für Rohdaten einer beliebigen Größe oder Form häufig ein Datensee: ein Speicher mit einer riesigen Datenmenge im nativen Format, ob strukturiert, halbstrukturiert oder unstrukturiert. Datenseen unterstützen moderne Anforderungen für die Big-Data-Analyse durch schnellere und flexiblere Datenaufnahme und -speicherung, damit Rohdaten auf vielfältige Weise schnell analysieren werden können.

Stream-Daten werden kontinuierlich von angeschlossenen Geräten und Apps generiert, die sich überall befinden, wie z. B. soziale Netzwerke, intelligente Zähler, Heimautomatisierung, Videospiele und IoT-Sensoren. Häufig werden diese Daten über Pipelines halbstrukturierter Daten gesammelt. Echtzeit-Analysen und Prognosealgorithmen können zwar auf Streams angewendet werden, jedoch werden die Streamdaten für die Analysen üblicherweise in Rohformaten mithilfe der Lambda-Architektur in einen Datensee wie Hadoop geleitet und dort gespeichert. Die Lambda-Architektur ist eine Datenverarbeitungsarchitektur, die entwickelt wurde, um riesige Datenmengen zu handhaben und dabei sowohl Batch- als auch Stream-Verarbeitungsmethoden zu nutzen. Das Design gleicht Latenz, Durchsatz und Fehlertoleranz aus. Es gibt heute eine Vielzahl von Optionen für das Streaming von Daten, darunter Amazon Kinesis, Storm, Flume, Kafka und Informatica Vibe Data Stream.



Datenseen bieten auch optimierte Verarbeitungsmechanismen über APIs oder SQL-ähnliche Sprachen für die Umwandlung von Rohdaten mit „Schema on Read“-Funktion. Sobald die Daten in einem Datensee gelandet sind, müssen sie aufgenommen und für die Analyse vorbereitet werden. Tableau hat Partner wie **Informatica**, **Alteryx**, **Trifacta** und Datameer, die beim Prozess helfen und reibungslos mit Tableau arbeiten. Wechselweise können Sie für die Selfservice-Datenvorbereitung **Tableau Prep** verwenden.

## Speicherung und Verarbeitung

Hadoop wird aufgrund seiner Widerstandsfähigkeit und der geringen Kosten, des ausskalierten Datenspeichers, der parallelen Verarbeitung und des geclusterten Arbeitsaufwandsmanagements für Datenseen verwendet. Hadoop wird oft als Big-Data-Plattform genutzt, ist aber keine Datenbank. Hadoop ist ein Open-Source-Software-Framework für die Speicherung von Daten und das Ausführen von Anwendungen auf Clustern von Standardhardware. Es bietet massiven Speicherplatz für jede Art von Daten, enorme Rechenleistung und die Fähigkeit, extreme Mengen an gleichzeitigen Aufgaben oder Aufträgen zu bewältigen.

In einer modernen Analysearchitektur bietet Hadoop kostengünstige Speicher- und Datenarchivierungslösungen für die Entladung alter historischer Daten aus dem Data Warehouse in Online-Kühlhäuser. Es wird auch bei Analysen in den Bereichen IoT, Datenwissenschaft und für unstrukturierte Analysen eingesetzt. Tableau bietet eine direkte Verbindung zu allen Hadoop-Hauptverteilern mit **Cloudera** über Impala, **Hortonworks** über Hive und **MapR** über Apache Drill.

Datenbanken und Data Warehouses werden in der modernen Analysearchitektur immer einen Platz haben. Sie spielen weiterhin eine entscheidende Rolle bei der Bereitstellung von kontrollierten, akkuraten und konformen Dimensionsdaten im gesamten Unternehmen für die Selfservice-Berichterstellung. Selbst Unternehmen, die andere Technologien einsetzen (z. B. Hadoop, Datenseen), behalten in der Regel relationale Datenbanken als Teil ihrer Datenquellenmischung. **Snowflake** ist ein Beispiel für ein Cloud-natives, SQL-basiertes Unternehmens-Data-Warehouse mit nativem Tableau-Konnektor.

Als Datenseen können auch Objektspeicher wie Simple Storage Service (S3) von Amazon Web Services und NoSQL-Datenbanken mit flexiblen Schemata verwendet werden. Tableau unterstützt den **Athena-Datendienst von Amazon** zur Verbindung mit Amazon S3 und verfügt über verschiedene Tools, die eine direkte Verbindung zu NoSQL-Datenbanken ermöglichen. Beispiele für NoSQL-Datenbanken, die häufig mit Tableau verwendet werden, sind u. a. **MongoDB**, **Datastax** und **MarkLogic**.

Die Datenwissenschafts- und Technikplattform **Databricks** ermöglicht eine Datenverarbeitung mit Spark, eine beliebte Funktionseinheit für sowohl batchorientierte als auch interaktive skalierte Datenverarbeitung. Sie können über einen nativen Connector zu Spark die Ergebnisse der komplexen Maschinenlernmodelle von Databricks in Tableau visualisieren.



## Abfragebeschleunigung

Für Big Data ist es möglich, das maschinelle Lernen zu nutzen und Sentiment-Analysen durchzuführen, aber als Erstes wird häufig die Frage gestellt: Wie schnell ist die interaktive SQL? SQL ist letztendlich die Datenleitung der Geschäftsanwender, die Big Data für schnellere, leichter wiederholbare KPI-Dashboards und Untersuchungsanalysen nutzen wollen.

Dieser Geschwindigkeitsbedarf hat die Aufnahme von schnelleren Datenbanken angetrieben und sich Technologien für eine speicherresidente und massive Parallelverarbeitung (MPP) zunutze gemacht, wie etwa **Exasol** und **MemSQL**, Hadoop-basierte Speicher wie Kudu und Technologien, die schnellere Abfragen durch Vorverarbeitung ermöglichen, wie z. B. **Vertica**. Diese Abfragebeschleuniger verwenden SQL-on-Hadoop-Engines wie Apache Impala, Hive LLAP, Presto, Phoenix und Drill sowie OLAP-on-Hadoop-Technologien wie **AtScale**, **Jethro Data** und Kyvos Insights und verwischen die Trennlinie zwischen herkömmlichen Data Warehouses und der Big-Data-Welt..

## Datenkatalog

Unternehmensdatenkataloge dienen allgemein als Geschäftsglossar mit Datenquellen und allgemeinen Datendefinitionen, die es Anwendern ermöglichen, leichter die richtigen Daten zur Entscheidungsfindung aus geregelten und genehmigten Datenquellen zu finden. Sie werden mit Metadaten aus Tabellen, Ansichten und gespeicherten Prozeduren gefüllt, indem einbezogene Datenquellen gescannt werden. Die Datenkuration kann sogar so weit gehen, Informationen und Web-Links aus Wissensdatenbanken aufzunehmen, um den Anwendern dabei zu helfen, den Kontext der Daten zu verstehen und eine intelligenterere Klassifizierung und automatisierte Datensuche zu ermöglichen.

Datenkataloge gibt es innerhalb von Visual-Analytics-Lösungen und sie sind auch als eigenständige Lösungen für die nahtlose Integration mit Tableau erhältlich. Unsere Datenkatalogpartner sind u. a. **Informatica**, **Alation**, **Unifi**, Collibra und Waterline.

## Big-Data-Analysearchitekturen

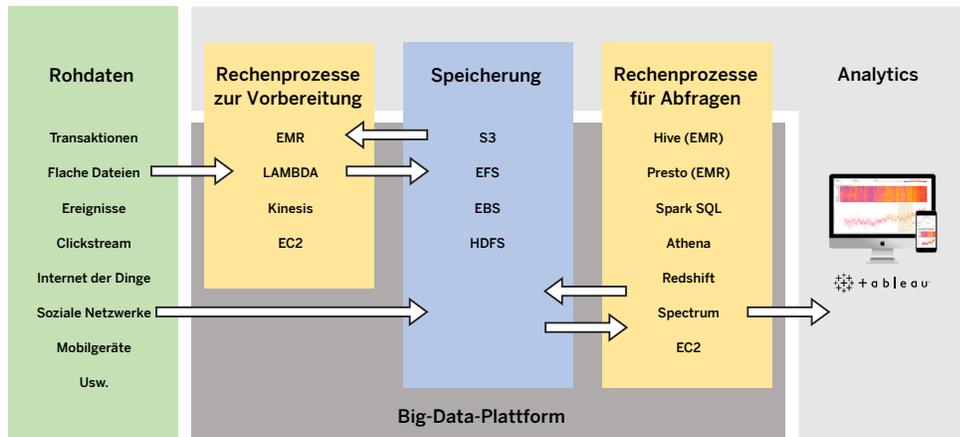
Man darf nicht vergessen, dass es in Bezug auf erfolgreiche Big-Data-Architekturen keine allgemeingültige Lösung gibt. Unsere Kunden haben für ihre Big-Data-Analysen einzigartige, zugeschnittene Lösungen mit verschiedenen Plattformen und Werkzeugen, um daraus ihre Datenpipelines zu erstellen. Dementsprechend werden wir zusätzliche Informationen zu geteilten Komponenten der Architekturen bereitstellen, die zum Erfolg dieser Big-Data-Analytics-Plattformen beigetragen haben.

**Haftungsausschluss:** Bitte beachten Sie, dass es sich bei den folgenden Beispielen um Tableau-Interpretationen handelt und diese nicht von den Cloud-Anbietern oder Kunden, die sie vertreten, erstellt wurden. Falls verfügbar, haben wir Links zu den ursprünglichen Darstellungen eingefügt. Bei diesen Grafiken handelt es sich um Verallgemeinerungen zur Verdeutlichung von Ähnlichkeiten bei Kernelementen unterschiedlicher Flows. Diese spiegeln möglicherweise nicht jeden Teil der vollständigen Big-Data-Analytics-Plattformen wider und stellen evtl. nur bestimmte Anwendungsfälle dar. Bedenken Sie außerdem, dass „Verrechnung zur Vorbereitung“ hauptsächlich „Prozess/Katalog“ repräsentiert, wo „Verrechnung für Abfrage“ hauptsächlich „Analysieren/Modell“ repräsentiert.

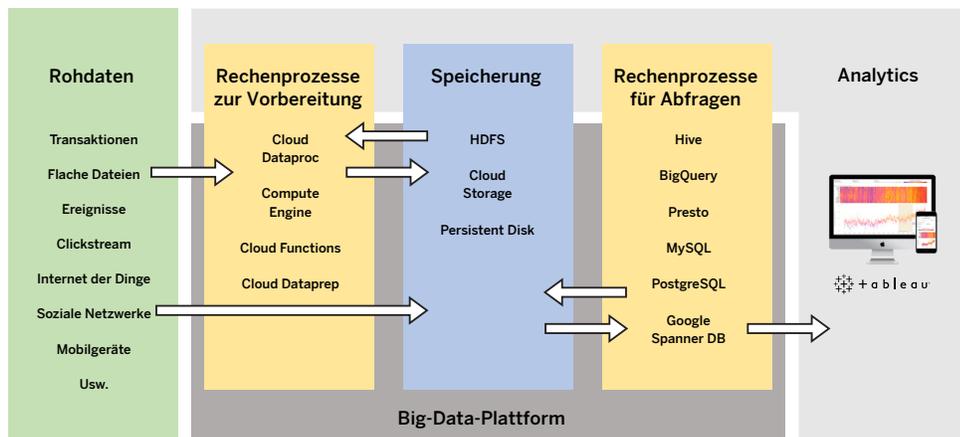


# Beispiele für die größten Cloud-Anbieter

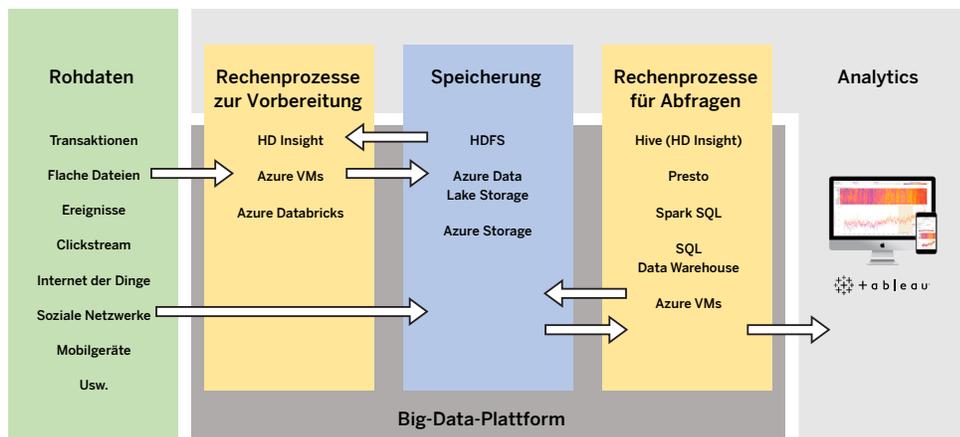
## Amazon Web Services



## Google Cloud Platform

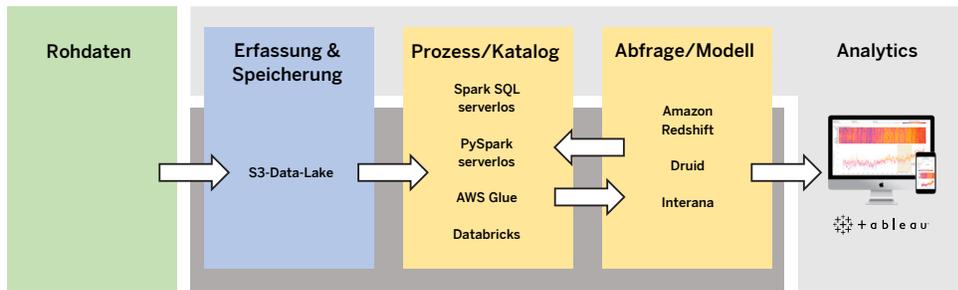


## Microsoft Azure

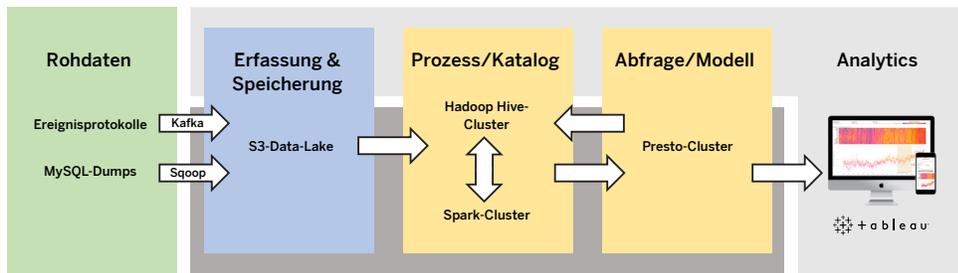


# Beispiele für Tableau-Kunden

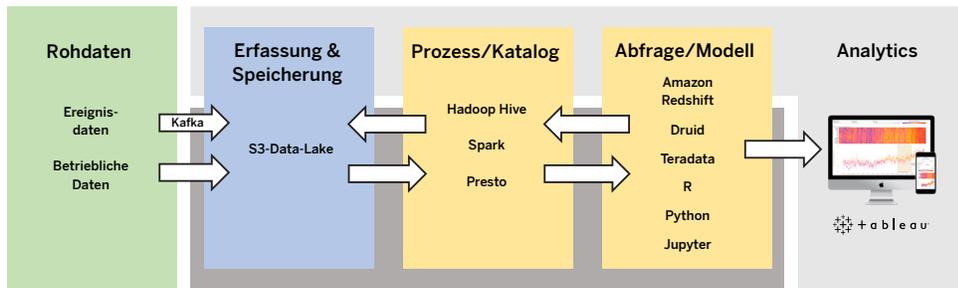
## Edmunds – mehr erfahren



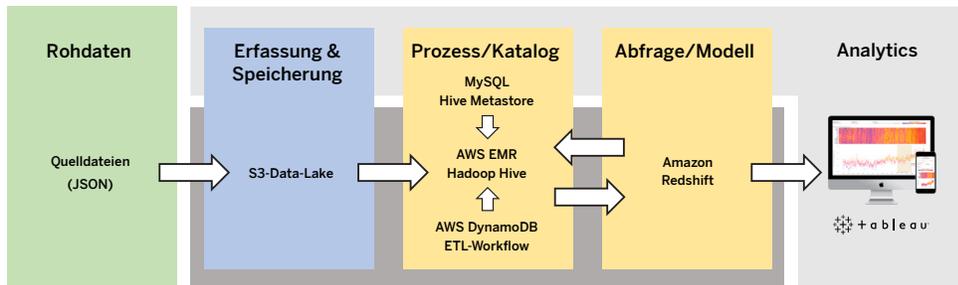
## Airbnb – mehr erfahren



## Netflix – mehr erfahren



## Expedia – mehr erfahren



## Gängige Muster

Obwohl keine Unternehmensarchitektur gleich einer anderen ist, kann das Erkennen ähnlicher Muster und deren Gemeinsamkeiten dabei helfen, eine Strategie für Ihre eigene Big-Data-Analytics-Plattform zu entwickeln. Folgendes haben wir bei erfolgreichen Big-Data-Analysearchitekturen durchgängig beobachtet:

- **Eine Speicherschicht** — Bei vielen bekannt als Speichersee. Ihre Datenstrategie benötigt möglicherweise mehrere Speicherumgebungen, sollte aber strukturierte, halbstrukturierte und unstrukturierte Daten enthalten.
- **Compute Engines mit und ohne Server** — Einige für die Datenvorbereitung und -analyse, andere Compute Engines für die Abfrage. Die dynamische Natur des serverlosen Computings ermöglicht mehr Flexibilität, da die Notwendigkeit einer vorherigen Ressourcenallokation entfällt.
- **Support für Datenvolumen, Bandbreite und Geschwindigkeit** — Dies gilt nicht nur für Daten, sondern auch für die wachsende Komplexität und Anzahl der Anwendungsfälle, die in Zukunft noch zunehmen wird.
- **Das richtige Werkzeug für den Auftrag** — Es ist wichtig, die Komponenten Ihrer Architektur entsprechend auf Ihre einzigartige Datenstrategie anzupassen, aber es ist genauso bedeutsam, vor dem Hintergrund sich ändernder Geschäftsbedürfnisse agil zu bleiben.
- **Führung und Sicherheit auf Unternehmensebene** — Obwohl wir diese Punkte nicht detailliert besprochen haben, handelt es sich bei der Sicherheit und Führung um Grundlagen, um die Skalierbarkeit und ordnungsgemäße Nutzung Ihrer Daten sicherzustellen.
- **Kostenbewusstsein** — Berücksichtigen Sie die Kosten, wenn Sie über die notwendige Leistung und Flexibilität für Ihre Big-Data-Architektur nachdenken. Die Cloud gewährt viel Elastizität für Wachstum, jedoch müssen Sie die finanziellen Auswirkungen Ihrer Datenspeicherung und -verarbeitung, Nebenläufigkeit, Latenz, Analyseanwendungsfälle usw. berücksichtigen.

Im Zuge der weiteren Entwicklung der Big-Data-Landschaft bleibt ein Aspekt bei allen Herausforderungen gleich: Unternehmen müssen in der Lage sein, eine gemeinsame, moderne Analyseplattform zu verwenden, um auf ihre Daten zuzugreifen, ob groß oder klein, und unabhängig vom Standort. Mit den richtigen Plattformen, Prozessen und Programmen zur Mitarbeiterförderung werden datengesteuerte Entscheidungen zu einem enormen Vorteil.



## Über Tableau

Tableau ist eine vollumfängliche, anwenderfreundliche, unternehmensfähige visuelle BI-Plattform, die den Benutzern hilft, ihre Daten zu sehen und verstehen – und zwar mithilfe von blitzschnellen Selfservice-Analytics im großen Maßstab. Egal ob Tableau auf lokalen Computern oder in der Cloud bzw. in einer Windows- oder Linux-Umgebung ausgeführt wird, die Lösung nutzt stets Ihre vorhandenen Technologie-Investitionen und lässt sich mühelos skalieren, wenn Ihre Datenumgebung verlagert wird und wächst. Optimale Nutzung Ihrer wertvollsten Ressourcen – Ihre Daten und Ihre Mitarbeiter

## Weitere Materialien

[Die Bausteine einer modernen Analytics-Plattform](#)

[Tableau Enterprise Analytics, powered by IT](#)

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

[Kostenlose Tableau-Testversion](#)