

AWS Leitfaden zur Entscheidungsfindung

# Auswahl eines AWS Analysedienstes



# Auswahl eines AWS Analysedienstes: AWS Leitfaden zur Entscheidungsfindung

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Die Handelsmarken und Handelsaufmachung von Amazon dürfen nicht in einer Weise in Verbindung mit nicht von Amazon stammenden Produkten oder Services verwendet werden, durch die Kunden irregeführt werden könnten oder Amazon in schlechtem Licht dargestellt oder diskreditiert werden könnte. Alle anderen Handelsmarken, die nicht Eigentum von Amazon sind, gehören den jeweiligen Besitzern, die möglicherweise zu Amazon gehören oder nicht, mit Amazon verbunden sind oder von Amazon gesponsert werden.

# Table of Contents

Leitfaden zur Entscheidungsfindung ..... 1

Einführung ..... 1

Verstehen ..... 2

Erwägen ..... 6

Klicken Sie auf ..... 15

Verwenden Sie ..... 20

Erkunden ..... 30

Dokumentverlauf ..... 32

..... xxxiii

# Auswahl eines AWS Analysedienstes

## Den ersten Schritt machen

Zweck	Finden Sie heraus, welche AWS Analysedienste für Ihr Unternehmen am besten geeignet sind.
Letzte Aktualisierung	24. September 2025
Abgedeckte Dienstleistungen	<ul style="list-style-type: none"><li>• <a href="#">Amazon Athena</a></li><li>• <a href="#">AWS Clean Rooms</a></li><li>• <a href="#">Amazon Data Firehose</a></li><li>• <a href="#">Amazon DataZone</a></li><li>• <a href="#">Amazon EMR</a></li><li>• <a href="#">AWS Glue</a></li><li>• <a href="#">Amazon Kinesis Data Streams</a></li><li>• <a href="#">Amazon Managed Service für Apache Flink</a></li><li>• <a href="#">Amazon Managed Streaming für Apache Kafka</a></li><li>• <a href="#">Amazon Managed Workflows für Apache Airflow</a></li><li>• <a href="#">OpenSearch Amazon-Dienst</a></li><li>• <a href="#">Schnelle Suite</a></li><li>• <a href="#">Amazon Redshift</a></li><li>• <a href="#">Amazon S3</a></li><li>• <a href="#">Amazon SageMaker</a></li></ul>

## Einführung

Daten sind für moderne Unternehmen von grundlegender Bedeutung. Menschen und Anwendungen müssen sicher auf Daten zugreifen und diese analysieren können, die aus neuen und vielfältigen Quellen stammen. Das Datenvolumen nimmt ebenfalls ständig zu, was dazu führen kann, dass

Unternehmen Schwierigkeiten haben, alle erforderlichen Daten zu erfassen, zu speichern und zu analysieren.

Um diesen Herausforderungen zu begegnen, müssen Sie eine moderne Datenarchitektur aufbauen, die all Ihre Datensilos für Analysen und Erkenntnisse — einschließlich Daten von Drittanbietern — aufbricht und sie für alle im Unternehmen an einem zentralen Ort zugänglich macht, sodass alle Mitarbeiter des Unternehmens auf diese Daten zugreifen können, und das alles unter Kontrolle. end-to-end Es wird auch immer wichtiger, Ihre Analytik- und Machine Learning-Systeme (ML) miteinander zu verbinden, um prädiktive Analysen zu ermöglichen.

Dieser Entscheidungsleitfaden hilft Ihnen dabei, die richtigen Fragen zu stellen, um Ihre moderne Datenarchitektur auf AWS Services aufzubauen. Es wird erklärt, wie Sie Ihre Datensilos (durch die Verbindung Ihres Data Lake und Ihrer Data Warehouses), Ihre Systemsilos (durch die Verbindung von ML und Analytics) und Ihre Personalsilos (indem Sie Daten in die Hände aller in Ihrem Unternehmen legen) aufbrechen können.

[Dieser achtminütige Auszug stammt aus einer einstündigen Präsentation von Irish Chandrasekaran und Rick Sears auf der re:Invent 2024. Es bietet einen Überblick darüber, wie ein fiktives Unternehmen, Maxdome, SageMaker Unified Studio KI und Analytics — Teil der nächsten Generation von Amazon SageMaker — nutzt, um Dateneinblicke zu gewinnen.](#)

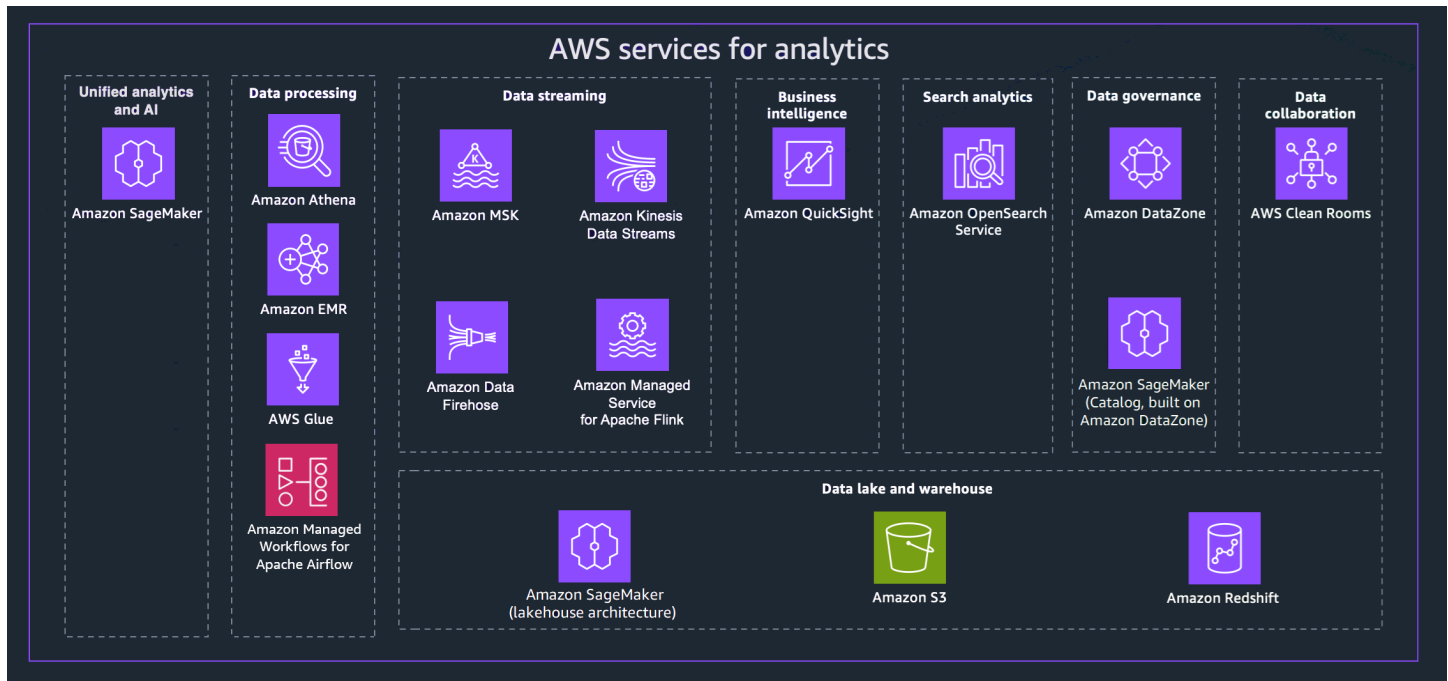
## Analyseservices verstehen AWS

Eine moderne Datenstrategie basiert auf einer Reihe von Technologiebausteinen, die Ihnen helfen, Daten zu verwalten, darauf zuzugreifen, sie zu analysieren und darauf zu reagieren. Sie bietet Ihnen auch mehrere Optionen, um eine Verbindung zu Datenquellen herzustellen. Eine moderne Datenstrategie sollte Ihre Teams in die Lage versetzen,

- Verwenden Sie Ihre bevorzugten Tools oder Techniken
- Verwenden Sie künstliche Intelligenz (KI), um Antworten auf spezifische Fragen zu Ihren Daten zu finden
- Steuern Sie mit den richtigen Sicherheits- und Datenverwaltungskontrollen, wer Zugriff auf Daten hat
- Brechen Sie Datensilos auf, um das Beste aus Data Lakes und speziell entwickelten Datenspeichern herauszuholen
- Speichern Sie beliebige Datenmengen kostengünstig und in offenen, standardbasierten Datenformaten

- Connect Sie Ihre Data Lakes, Data Warehouses, Betriebsdatenbanken, Anwendungen und föderierten Datenquellen zu einem kohärenten Ganzen

AWS bietet eine Vielzahl von Services, die Sie bei der Umsetzung einer modernen Datenstrategie unterstützen. Das folgende Diagramm zeigt die AWS Analysedienste, die in diesem Handbuch behandelt werden. Die folgenden Registerkarten enthalten zusätzliche Details.



## Unified analytics and AI

Die nächste Generation von [Amazon SageMaker](#) kombiniert weit verbreitete Funktionen für AWS maschinelles Lernen (ML) und Analysen, um ein integriertes Erlebnis für Analytik und KI zu bieten und einen einheitlichen Zugriff auf all Ihre Daten zu ermöglichen. Mit [Amazon SageMaker Unified Studio](#) können Sie mit vertrauten AWS Tools für Modellentwicklung, generative KI-Anwendungsentwicklung, Datenverarbeitung und SQL-Analyse schneller zusammenarbeiten und schneller arbeiten. All dies wird durch Amazon Q Developer, unseren generativen KI-Assistenten für die Softwareentwicklung, beschleunigt. Greifen Sie über Data Lakes, Data Warehouses oder Drittanbieter- und Verbundquellen auf Ihre Daten zu — mit integrierter Governance, um die Sicherheitsanforderungen Ihres Unternehmens zu erfüllen.

## Data processing

- [Amazon Athena](#) hilft Ihnen bei der Analyse unstrukturierter, halbstrukturierter und strukturierter Daten, die in Amazon S3 gespeichert sind. Beispiele hierfür sind CSV und JSON oder

spaltenbasierte Datenformate wie Apache Parquet und Apache ORC. Mit Athena lassen sich Ad-hoc-Abfragen über ANSI SQL ausführen; dabei müssen die Daten weder aggregiert noch in Athena geladen werden. Athena lässt sich in Quick Suite und [andere AWS Dienste](#) integrieren. AWS Glue Data Catalog Sie können mit [Trino](#) auch Daten in großem Umfang analysieren, ohne die Infrastruktur verwalten zu müssen, und mithilfe von Apache Flink und Apache Spark Echtzeitanalysen erstellen.

- [Amazon EMR](#) ist eine verwaltete Cluster-Plattform, die die Ausführung von Big-Data-Frameworks wie Apache Hadoop und Apache Spark vereinfacht, AWS um riesige Datenmengen zu verarbeiten und zu analysieren. Die Verwendung dieser Frameworks und verwandter Open-Source-Projekte, können Sie Daten zu Analysezielen und Business-Intelligence-Workloads verarbeiten. Mit Amazon EMR können Sie auch große Datenmengen in und aus anderen AWS Datenspeichern und Datenbanken wie Amazon S3 transformieren und verschieben.
- Mit [AWS Glue](#) können Sie mehr als 100 verschiedene Datenquellen entdecken und eine Verbindung zu ihnen herstellen und Ihre Daten in einem zentralen Datenkatalog verwalten. Sie können ETL-Pipelines visuell erstellen, ausführen und überwachen, um Daten in Ihre Data Lakes zu laden. Außerdem können Sie mit Athena, Amazon EMR und Amazon Redshift Spectrum sofort katalogisierte Daten suchen und abfragen.
- [Amazon Managed Workflows for Apache Airflow](#) (MWAA) ist eine vollständig verwaltete Implementierung von Apache Airflow, die es einfacher macht, Datenworkflows in der Cloud zu erstellen, zu planen und zu überwachen. MWAA passt die Workflow-Kapazität automatisch an Ihre Bedürfnisse an und lässt sich in Sicherheitsdienste integrieren. AWS Sie können MWAA verwenden, um Workflows für alle Ihre Analysedienste zu orchestrieren, einschließlich Datenverarbeitung, ETL-Jobs und Pipelines für maschinelles Lernen.

## Data streaming

- Mit [Amazon Managed Streaming for Apache Kafka](#) (Amazon MSK) können Sie Anwendungen erstellen und ausführen, die Apache Kafka zur Verarbeitung von Streaming-Daten verwenden. Amazon MSK stellt die Vorgänge auf Steuerebene bereit, z. B. zum Erstellen, Aktualisieren und Löschen von Clustern. Damit können Sie Apache Kafka-Operationen auf Datenebene verwenden, z. B. zum Erstellen und Nutzen von Daten.
- Mit [Amazon Kinesis Data Streams](#) können Sie große Datenströme in Echtzeit sammeln und verarbeiten. Der verwendete Datentyp kann Protokolldaten zur IT-Infrastruktur, Anwendungsprotokolle, Data-Feeds von sozialen Medien, Marktdaten-Feeds sowie Web-Clickstream-Daten einschließen.

- [Amazon Data Firehose](#) ist ein vollständig verwalteter Service für die Bereitstellung von Echtzeit-Streaming-Daten an Ziele wie Amazon S3, Amazon Redshift, Amazon OpenSearch Service, Splunk und Apache Iceberg Tables. Sie können auch Daten an jeden benutzerdefinierten HTTP-Endpunkt oder an HTTP-Endpunkte senden, deren Eigentümer unterstützte Drittanbieter sind, darunter Datadog, Dynatrace, MongoDB, New Relic LogicMonitor, Coralogix und Elastic.
- Mit [Amazon Managed Service für Apache Flink](#) können Sie Java, Scala, Python oder SQL verwenden, um Streaming-Daten zu verarbeiten und zu analysieren. Sie können Code für Streaming-Quellen und statische Quellen erstellen und ausführen, um Zeitreihenanalysen durchzuführen, Echtzeit-Dashboards und Metriken bereitzustellen.

## Business intelligence

[Quick Suite](#) bietet Entscheidungsträgern die Möglichkeit, Informationen in einer interaktiven visuellen Umgebung zu untersuchen und zu interpretieren. In einem einzigen Daten-Dashboard kann Quick Suite AWS Daten, Daten von Drittanbietern, Big Data, Tabellenkalkulationsdaten, SaaS-Daten, B2B-Daten und mehr enthalten. Mit Quick Suite Q können Sie natürliche Sprache verwenden, um Fragen zu Ihren Daten zu stellen und eine Antwort zu erhalten. Zum Beispiel: „Was sind die meistverkauften Kategorien in Kalifornien?“

## Search analytics

[Amazon OpenSearch Service](#) stellt alle Ressourcen für Ihren OpenSearch Cluster bereit und startet ihn. Außerdem werden ausgefallene OpenSearch Serviceknoten automatisch erkannt und ersetzt, wodurch der mit selbstverwalteten Infrastrukturen verbundene Aufwand reduziert wird. Sie können OpenSearch Service Direct Query verwenden, um Daten in Amazon S3 und anderen AWS Diensten zu analysieren.

## Data governance

Mit [Amazon DataZone](#) können Sie den Zugriff auf Daten mithilfe detaillierter Kontrollen verwalten und steuern. Diese Kontrollen tragen dazu bei, den Zugriff mit den richtigen Rechten und dem richtigen Kontext sicherzustellen. Amazon DataZone vereinfacht Ihre Architektur durch die Integration von Datenverwaltungsdiensten wie Amazon Redshift, Athena, Quick Suite AWS Glue, lokalen Quellen und Quellen von Drittanbietern.

## Data collaboration

[AWS Clean Rooms](#) ist ein sicherer Arbeitsbereich für die Zusammenarbeit, in dem Sie kollektive Datensätze analysieren können, ohne Zugriff auf die Rohdaten zu gewähren. Sie können mit anderen Unternehmen zusammenarbeiten, indem Sie die Partner auswählen, mit denen Sie



zusammenarbeiten möchten, deren Datensätze auswählen und Kontrollen zur Verbesserung der Privatsphäre für diese Partner konfigurieren. Wenn Sie Abfragen ausführen, AWS Clean Rooms liest es Daten vom ursprünglichen Speicherort dieser Daten und wendet integrierte Analyseregeln an, damit Sie die Kontrolle über diese Daten behalten.

## Data lake and data warehouse

- [Die nächste Generation von Amazon SageMaker](#) ist vollständig mit Apache Iceberg kompatibel, sodass Sie Daten aus Amazon Simple Storage Service (Amazon S3) -Data Lakes und Amazon Redshift Data Warehouses vereinheitlichen können. Dies ermöglicht die Erstellung von Analyse-, KI- und ML-Anwendungen (Machine Learning) auf einer einzigen Datenkopie. Durch Zero-ETL-Integrationen können Sie Daten aus operativen Quellen nahezu in Echtzeit streamen, föderierte Abfragen über mehrere Quellen hinweg ausführen und mit Apache Iceberg-kompatiblen Tools auf Daten zugreifen. Sie können Ihre Daten schützen, indem Sie detaillierte Berechtigungen definieren, die für alle Ihre Analyse- und ML-Tools und Engines gelten.
- [Amazon S3](#) kann praktisch jede Menge und Art von Daten speichern und schützen, die Sie als Grundlage für Ihren Data Lake verwenden können. Amazon S3 bietet Verwaltungsfunktionen, mit denen Sie den Zugriff auf Ihre Daten optimieren, organisieren und konfigurieren können, um Ihre spezifischen geschäftlichen, organisatorischen und Compliance-Anforderungen zu erfüllen. Amazon S3 S3-Tabellen bieten S3-Speicher, der für Analyse-Workloads optimiert ist. Mithilfe von Standard-SQL-Anweisungen können Sie Ihre Tabellen mit Abfrage-Engines abfragen, die Iceberg unterstützen, wie Athena, Amazon Redshift und Apache Spark.
- [Amazon Redshift](#) ist ein vollständig verwalteter Data-Warehouse-Service im Petabyte-Bereich. Amazon Redshift kann mit einem Data Lakehouse in Amazon verbunden werden SageMaker, sodass Sie die leistungsstarken SQL-Analysefunktionen für Ihre vereinheitlichten Daten in Amazon Redshift Redshift-Data Warehouses und Amazon S3 S3-Data Lakes nutzen können. Sie können Amazon Q auch in Amazon Redshift verwenden, was die SQL-Erstellung durch natürliche Sprache vereinfacht.

## Berücksichtigen Sie die Kriterien für Analysedienste AWS

Es gibt viele Gründe, auf Datenanalysen aufzubauen AWS. Möglicherweise müssen Sie als ersten Schritt auf Ihrer Cloud-Migration ein Projekt auf der grünen Wiese oder ein Pilotprojekt unterstützen. Alternativ können Sie einen vorhandenen Workload mit so wenig Unterbrechungen wie möglich migrieren. Was auch immer Ihr Ziel ist, die folgenden Überlegungen können Ihnen dabei helfen, Ihre Wahl zu treffen.

## Assess data sources and data types

Analysieren Sie die verfügbaren Datenquellen und Datentypen, um ein umfassendes Verständnis der Datenvielfalt, Häufigkeit und Qualität zu erlangen. Machen Sie sich mit potenziellen Herausforderungen bei der Verarbeitung und Analyse der Daten vertraut. Diese Analyse ist aus folgenden Gründen von entscheidender Bedeutung:

- Datenquellen sind vielfältig und stammen aus verschiedenen Systemen, Anwendungen, Geräten und externen Plattformen.
- Datenquellen haben eine einzigartige Struktur, ein einzigartiges Format und eine einzigartige Häufigkeit von Datenaktualisierungen. Die Analyse dieser Quellen hilft bei der Identifizierung geeigneter Methoden und Technologien zur Datenerfassung.
- Durch die Analyse von Datentypen wie strukturierten, halbstrukturierten und unstrukturierten Daten werden die geeigneten Datenverarbeitungs- und Speicheransätze bestimmt.
- Die Analyse von Datenquellen und -typen erleichtert die Bewertung der Datenqualität und hilft Ihnen, potenzielle Probleme mit der Datenqualität — fehlende Werte, Inkonsistenzen oder Ungenauigkeiten — zu antizipieren.

## Data processing requirements

Ermitteln Sie die Datenverarbeitungsanforderungen für die Art und Weise, wie Daten aufgenommen, transformiert, bereinigt und für die Analyse vorbereitet werden. Zu den wichtigsten Überlegungen gehören:

- **Datentransformation:** Ermitteln Sie die spezifischen Transformationen, die erforderlich sind, um die Rohdaten für die Analyse geeignet zu machen. Dies beinhaltet Aufgaben wie Datenaggregation, Normalisierung, Filterung und Anreicherung.
- **Datenbereinigung:** Beurteilen Sie die Datenqualität und definieren Sie Prozesse für den Umgang mit fehlenden, ungenauen oder inkonsistenten Daten. Implementieren Sie Datenbereinigungstechniken, um qualitativ hochwertige Daten für zuverlässige Erkenntnisse zu gewährleisten.
- **Verarbeitungshäufigkeit:** Ermitteln Sie anhand der Analyseanforderungen, ob eine Verarbeitung in Echtzeit, nahezu in Echtzeit oder Batch-Verarbeitung erforderlich ist. Die Verarbeitung in Echtzeit ermöglicht sofortige Erkenntnisse, während die Stapelverarbeitung für regelmäßige Analysen ausreichend sein kann.
- **Skalierbarkeit und Durchsatz:** Bewerten Sie die Skalierbarkeitsanforderungen für den Umgang mit Datenmengen, die Verarbeitungsgeschwindigkeit und die Anzahl gleichzeitiger

Datenanfragen. Stellen Sie sicher, dass der gewählte Verarbeitungsansatz dem future Wachstum Rechnung trägt.

- Latenz: Berücksichtigen Sie die akzeptable Latenz für die Datenverarbeitung und die Zeit, die von der Datenaufnahme bis zu den Analyseergebnissen benötigt wird. Dies ist besonders wichtig für Echtzeit- oder zeitkritische Analysen.

## Storage requirements

Ermitteln Sie den Speicherbedarf, indem Sie festlegen, wie und wo Daten in der Analysepipeline gespeichert werden. Zu den wichtigen Überlegungen gehören:

- Datenvolumen: Beurteilen Sie die Menge der generierten und gesammelten Daten und schätzen Sie das future Datenwachstum ab, um ausreichend Speicherkapazität zu planen.
- Datenspeicherung: Definieren Sie die Dauer, für die Daten für historische Analysen oder Compliance-Zwecke aufbewahrt werden sollen. Ermitteln Sie die geeigneten Richtlinien zur Datenspeicherung.
- Datenzugriffsmuster: Erfahren Sie, wie auf Daten zugegriffen und wie Daten abgefragt werden, um die am besten geeignete Speicherlösung auszuwählen. Berücksichtigen Sie Lese- und Schreibvorgänge, Datenzugriffshäufigkeit und Datenlokalität.
- Datensicherheit: Priorisieren Sie die Datensicherheit, indem Sie Verschlüsselungsoptionen, Zugriffskontrollen und Datenschutzmechanismen zum Schutz vertraulicher Informationen evaluieren.
- Kostenoptimierung: Optimieren Sie die Speicherkosten, indem Sie die kostengünstigsten Speicherlösungen auf der Grundlage von Datenzugriffsmustern und -nutzung auswählen.
- Integration mit Analysediensten: Sorgen Sie für eine nahtlose Integration zwischen der ausgewählten Speicherlösung und den in der Pipeline befindlichen Datenverarbeitungs- und Analysetools.

## Types of data

Wenn Sie sich für Analysedienste für die Erfassung und Erfassung von Daten entscheiden, sollten Sie verschiedene Arten von Daten berücksichtigen, die für die Bedürfnisse und Ziele Ihres Unternehmens relevant sind. Zu den gängigen Datentypen, die Sie möglicherweise berücksichtigen müssen, gehören:

- **Transaktionsdaten:** Dazu gehören Informationen über einzelne Interaktionen oder Transaktionen, wie Kundenkäufe, Finanztransaktionen, Online-Bestellungen und Benutzeraktivitätsprotokolle.
- **Dateibasierte Daten:** Bezieht sich auf strukturierte oder unstrukturierte Daten, die in Dateien wie Protokolldateien, Tabellen, Dokumenten, Bildern, Audiodateien und Videodateien gespeichert sind. Analysedienste sollten die Aufnahme verschiedener Dateiformate unterstützen.
- **Ereignisdaten:** Erfasst wichtige Ereignisse oder Vorfälle, wie Benutzeraktionen, Systemereignisse, Computerereignisse oder Geschäftsereignisse. Zu Ereignissen können alle Daten gehören, die mit hoher Geschwindigkeit eintreffen und für die Verarbeitung während oder nach der Verarbeitung erfasst werden.

## Operational considerations

Sie und teilen sich die operative Verantwortung AWS, wobei die Aufteilung der Verantwortung je nach Modernisierungsgrad unterschiedlich ist. Sie haben die Möglichkeit, Ihre Analyseinfrastruktur selbst zu verwalten AWS oder die zahlreichen serverlosen Analysedienste zu nutzen, um den Aufwand für das Infrastrukturmanagement zu verringern.

Selbstverwaltete Optionen geben Benutzern mehr Kontrolle über die Infrastruktur und die Konfigurationen, erfordern jedoch mehr betrieblichen Aufwand.

Serverlose Optionen reduzieren einen Großteil des betrieblichen Aufwands und bieten automatische Skalierbarkeit, hohe Verfügbarkeit und robuste Sicherheitsfunktionen, sodass sich Benutzer mehr auf die Entwicklung analytischer Lösungen und die Gewinnung von Erkenntnissen konzentrieren können, anstatt Infrastruktur- und Betriebsaufgaben zu verwalten. Beachten Sie die folgenden Vorteile serverloser Analyselösungen:

- **Abstraktion der Infrastruktur:** Serverlose Dienste abstrahieren das Infrastrukturmanagement und entlasten die Benutzer von Bereitstellungs-, Skalierungs- und Wartungsaufgaben. AWS kümmert sich um diese betrieblichen Aspekte und reduziert so den Verwaltungsaufwand.
- **Automatische Skalierung und Leistung:** Serverlose Dienste skalieren Ressourcen automatisch auf der Grundlage der Workload-Anforderungen und gewährleisten so eine optimale Leistung ohne manuelles Eingreifen.
- **Hochverfügbarkeit und Notfallwiederherstellung:** AWS bietet hohe Verfügbarkeit für serverlose Dienste. AWS verwaltet Datenredundanz, Replikation und Notfallwiederherstellung, um die Datenverfügbarkeit und Zuverlässigkeit zu verbessern.

- **Sicherheit und Compliance:** AWS verwaltet Sicherheitsmaßnahmen, Datenverschlüsselung und Compliance für serverlose Dienste und hält sich dabei an Industriestandards und bewährte Verfahren.
- **Überwachung und Protokollierung:** AWS bietet integrierte Überwachungs-, Protokollierungs- und Warnfunktionen für serverlose Dienste. Benutzer können über Amazon auf detaillierte Metriken und Protokolle zugreifen CloudWatch.

## Type of workload

Beim Aufbau einer modernen Analysepipeline ist es entscheidend, zu entscheiden, welche Arten von Workloads unterstützt werden sollen, um unterschiedlichen Analyseanforderungen effektiv gerecht zu werden. Zu den wichtigsten Entscheidungspunkten, die für jede Art von Workload berücksichtigt werden müssen, gehören:

### Batch-Arbeitslast

- **Datenvolumen und Häufigkeit:** Die Stapelverarbeitung eignet sich für große Datenmengen mit regelmäßigen Aktualisierungen.
- **Datenlatenz:** Die Batch-Verarbeitung kann im Vergleich zur Echtzeitverarbeitung zu Verzögerungen bei der Bereitstellung von Erkenntnissen führen.

### Interaktive Analyse

- **Komplexität der Datenabfrage:** Interaktive Analysen erfordern Antworten mit geringer Latenz für schnelles Feedback.
- **Datenvisualisierung:** Evaluieren Sie den Bedarf an interaktiven Tools zur Datenvisualisierung, damit Geschäftsanwender Daten visuell untersuchen können.

### Workloads streamen

- **Datengeschwindigkeit und Datenvolumen:** Streaming-Workloads erfordern eine Verarbeitung in Echtzeit, um Hochgeschwindigkeitsdaten verarbeiten zu können.
- **Datenfenstergestaltung:** Definieren Sie Datenfenstergestaltung und zeitbasierte Aggregationen für Streaming-Daten, um relevante Erkenntnisse zu gewinnen.

## Type of analysis needed

Definieren Sie klar die Geschäftsziele und die Erkenntnisse, die Sie aus den Analysen gewinnen möchten. Verschiedene Arten von Analysen dienen unterschiedlichen Zwecken. Zum Beispiel:

- Deskriptive Analysen sind ideal, um sich einen historischen Überblick zu verschaffen
- Diagnostische Analysen helfen dabei, die Gründe für vergangene Ereignisse zu verstehen
- Predictive Analytics prognostiziert future Ergebnisse
- Präskriptive Analysen bieten Empfehlungen für optimale Maßnahmen

Ordnen Sie Ihre Geschäftsziele den relevanten Analysetypen zu. Hier sind einige wichtige Entscheidungskriterien, die Ihnen bei der Auswahl der richtigen Analysetypen helfen sollen:

- Datenverfügbarkeit und Qualität: Deskriptive und diagnostische Analysen basieren auf historischen Daten, während prädiktive und präskriptive Analysen ausreichend historische Daten und hochwertige Daten erfordern, um genaue Modelle zu erstellen.
- Datenvolumen und Komplexität: Prädiktive und präskriptive Analysen erfordern umfangreiche Datenverarbeitungs- und Rechenressourcen. Stellen Sie sicher, dass Ihre Infrastruktur und Tools das Datenvolumen und die Komplexität bewältigen können.
- Komplexität der Entscheidungen: Wenn Entscheidungen mehrere Variablen, Einschränkungen und Ziele beinhalten, eignen sich präskriptive Analysen möglicherweise besser als Richtschnur für optimale Maßnahmen.
- Risikotoleranz: Präskriptive Analysen können zwar Empfehlungen geben, sind aber mit Unsicherheiten behaftet. Stellen Sie sicher, dass die Entscheidungsträger die mit den Analyseergebnissen verbundenen Risiken verstehen.

## Evaluate scalability and performance

Beurteilen Sie die Skalierbarkeits- und Leistungsanforderungen der Architektur. Das Design muss steigenden Datenmengen, Benutzeranforderungen und analytischen Workloads gerecht werden. Zu den wichtigsten Entscheidungsfaktoren, die es zu berücksichtigen gilt, gehören:

- Datenvolumen und Wachstum: Beurteilen Sie das aktuelle Datenvolumen und antizipieren Sie future Wachstum.
- Datengeschwindigkeit und Echtzeitanforderungen: Ermitteln Sie, ob die Daten in Echtzeit oder nahezu in Echtzeit verarbeitet und analysiert werden müssen.

- **Komplexität der Datenverarbeitung:** Analysieren Sie die Komplexität Ihrer Datenverarbeitungs- und Analyseaufgaben. Für rechenintensive Aufgaben bieten Dienste wie Amazon EMR eine skalierbare und verwaltete Umgebung für die Verarbeitung großer Datenmengen.
- **Parallelität und Benutzerauslastung:** Berücksichtigen Sie die Anzahl der gleichzeitigen Benutzer und den Grad der Benutzerlast auf dem System.
- **Funktionen zur auto-scaling:** Ziehen Sie Dienste in Betracht, die Funktionen zur automatischen Skalierung bieten, sodass Ressourcen je nach Bedarf automatisch nach oben oder unten skaliert werden können. Dies gewährleistet eine effiziente Ressourcennutzung und Kostenoptimierung.
- **Geografische Verteilung:** Ziehen Sie Dienste mit globaler Replikation und Datenzugriff mit geringer Latenz in Betracht, wenn Ihre Datenarchitektur über mehrere Regionen oder Standorte verteilt werden muss.
- **Kompromiss zwischen Kosten und Leistung:** Sorgen Sie für ein ausgewogenes Verhältnis zwischen Leistungsanforderungen und Kostenaspekten. Dienste mit hoher Leistung können mit höheren Kosten verbunden sein.
- **Service Level Agreements (SLAs):** Überprüfen Sie, ob die von SLAs Ihnen bereitgestellten AWS Dienste Ihren Erwartungen an Skalierbarkeit und Leistung entsprechen.

## Data governance

Bei Data Governance handelt es sich um eine Reihe von Prozessen, Richtlinien und Kontrollen, die Sie implementieren müssen, um eine effektive Verwaltung, Qualität, Sicherheit und Konformität Ihrer Datenbestände zu gewährleisten. Zu den wichtigsten Entscheidungspunkten, die Sie berücksichtigen sollten, gehören:

- **Richtlinien zur Datenspeicherung:** Definieren Sie Richtlinien zur Datenspeicherung auf der Grundlage gesetzlicher Anforderungen und Geschäftsanforderungen und richten Sie Prozesse für die sichere Entsorgung von Daten ein, wenn sie nicht mehr benötigt werden.
- **Prüfpfad und Protokollierung:** Entscheiden Sie sich für die Protokollierungs- und Prüfmechanismen zur Überwachung des Datenzugriffs und der Datennutzung. Implementieren Sie umfassende Prüfprotokolle, um Datenänderungen, Zugriffsversuche und Benutzeraktivitäten nachzuverfolgen, um die Einhaltung von Vorschriften und die Sicherheit zu überwachen.
- **Compliance-Anforderungen:** Machen Sie sich mit den branchenspezifischen und geografischen Datenschutzbestimmungen vertraut, die für Ihr Unternehmen gelten. Stellen Sie sicher, dass die Datenarchitektur diesen Vorschriften und Richtlinien entspricht.

- **Datenklassifizierung:** Klassifizieren Sie Daten anhand ihrer Sensibilität und definieren Sie geeignete Sicherheitskontrollen für jede Datenklasse.
- **Notfallwiederherstellung und Geschäftskontinuität:** Planen Sie Notfallwiederherstellung und Geschäftskontinuität, um die Datenverfügbarkeit und Widerstandsfähigkeit bei unerwarteten Ereignissen oder Systemausfällen sicherzustellen.
- **Datenaustausch durch Dritte:** Wenn Sie Daten an Dritte weitergeben, sollten Sie sichere Protokolle und Vereinbarungen für den Datenaustausch einführen, um die Vertraulichkeit der Daten zu schützen und Datenmissbrauch zu verhindern.

## Security

Die Sicherheit der Daten in der Analysepipeline beinhaltet den Schutz der Daten in jeder Phase der Pipeline, um deren Vertraulichkeit, Integrität und Verfügbarkeit zu gewährleisten. Zu den wichtigsten Entscheidungspunkten, die es zu berücksichtigen gilt, gehören:

- **Zugriffskontrolle und Autorisierung:** Implementieren Sie robuste Authentifizierungs- und Autorisierungsprotokolle, um sicherzustellen, dass nur autorisierte Benutzer auf bestimmte Datenressourcen zugreifen können.
- **Datenverschlüsselung:** Wählen Sie geeignete Verschlüsselungsmethoden für Daten, die in Datenbanken und Data Lakes gespeichert sind, und für die Übertragung von Daten zwischen verschiedenen Komponenten der Architektur.
- **Datenmaskierung und Anonymisierung:** Berücksichtigen Sie die Notwendigkeit einer Datenmaskierung oder Anonymisierung, um sensible Daten wie personenbezogene Daten oder sensible Geschäftsdaten zu schützen und gleichzeitig die Fortsetzung bestimmter Analyseprozesse zu ermöglichen.
- **Sichere Datenintegration:** Etablieren Sie sichere Datenintegrationsverfahren, um sicherzustellen, dass Daten sicher zwischen verschiedenen Komponenten der Architektur fließen, sodass Datenlecks oder unbefugter Zugriff während der Datenbewegung vermieden werden.
- **Netzwerkisolierung:** Ziehen Sie Dienste in Betracht, die [Amazon VPC-Endpunkte](#) unterstützen, um zu verhindern, dass Ressourcen dem öffentlichen Internet ausgesetzt werden.

## Plan for integration and data flows

Definieren Sie die Integrationspunkte und Datenflüsse zwischen den verschiedenen Komponenten der Analysepipeline, um einen reibungslosen Datenfluss und Interoperabilität



sicherzustellen. Zu den wichtigsten Entscheidungspunkten, die es zu berücksichtigen gilt, gehören:

- **Datenquellenintegration:** Identifizieren Sie die Datenquellen, aus denen Daten gesammelt werden, z. B. Datenbanken, Anwendungen, Dateien oder externe Daten APIs. Entscheiden Sie sich für die Datenaufnahmefunktionen (Batch, Echtzeit, ereignisbasiert), um Daten effizient und mit minimaler Latenz in die Pipeline zu bringen.
- **Datentransformation:** Ermitteln Sie die Transformationen, die zur Vorbereitung der Daten für die Analyse erforderlich sind. Entscheiden Sie sich für die Tools und Prozesse, mit denen die Daten bereinigt, aggregiert, normalisiert oder angereichert werden sollen, während sie die Pipeline durchlaufen.
- **Architektur für die Datenverlagerung:** Wählen Sie die geeignete Architektur für die Datenbewegung zwischen Pipeline-Komponenten. Ziehen Sie je nach den Echtzeitanforderungen und dem Datenvolumen eine Batch-Verarbeitung, Stream-Verarbeitung oder eine Kombination aus beidem in Betracht.
- **Datenreplikation und -synchronisierung:** Entscheiden Sie sich für Datenreplikations- und Synchronisationsmechanismen, um die Daten up-to-date über alle Komponenten hinweg zu speichern. Ziehen Sie je nach den Anforderungen an die Datenaktualität Replikationslösungen in Echtzeit oder regelmäßige Datensynchronisierungen in Betracht.
- **Datenqualität und Validierung:** Implementieren Sie Datenqualitätsprüfungen und Validierungsschritte, um die Integrität der Daten während des Transports durch die Pipeline sicherzustellen. Entscheiden Sie, welche Maßnahmen ergriffen werden sollen, wenn die Datenvalidierung fehlschlägt, z. B. Warnmeldungen oder Fehlerbehandlung.
- **Datensicherheit und Verschlüsselung:** Legen Sie fest, wie Daten während der Übertragung und im Speicher geschützt werden. Entscheiden Sie sich für die Verschlüsselungsmethoden zum Schutz sensibler Daten in der gesamten Pipeline und berücksichtigen Sie dabei das Sicherheitsniveau, das je nach Datensensibilität erforderlich ist.
- **Skalierbarkeit und Belastbarkeit:** Stellen Sie sicher, dass das Datenflussdesign horizontale Skalierbarkeit ermöglicht und größere Datenmengen und erhöhten Datenverkehr bewältigen kann.

## Architect for cost optimization

Der Aufbau Ihrer Analysepipeline AWS bietet verschiedene Möglichkeiten zur Kostenoptimierung. Um die Kosteneffizienz zu gewährleisten, sollten Sie die folgenden Strategien in Betracht ziehen:

- **Dimensionierung und Auswahl von Ressourcen:** Richten Sie die Größe Ihrer Ressourcen auf der Grundlage der tatsächlichen Arbeitslastanforderungen aus. Wählen Sie AWS Dienste und Instanztypen, die den Leistungsanforderungen der Workloads entsprechen und gleichzeitig eine Überprovisionierung vermeiden.
- **auto-scaling:** Implementieren Sie automatische Skalierung für Dienste, die unterschiedlichen Workloads ausgesetzt sind. Die automatische Skalierung passt die Anzahl der Instanzen dynamisch an den Bedarf an und senkt so die Kosten in Zeiten mit geringem Datenverkehr.
- **Spot-Instances:** Verwenden Sie Amazon EC2 Spot-Instances für unkritische und fehlertolerante Workloads. Spot-Instances können die Kosten im Vergleich zu On-Demand-Instances erheblich senken.
- **Reserved Instances:** Erwägen Sie den Kauf von AWS Reserved Instances, um erhebliche Kosteneinsparungen gegenüber On-Demand-Preisen für stabile Workloads mit vorhersehbarer Nutzung zu erzielen.
- **Datenspeicher-Tiering:** Optimieren Sie die Datenspeicherkosten, indem Sie je nach Datenzugriffshäufigkeit unterschiedliche Speicherklassen verwenden.
- **Richtlinien für den Datenlebenszyklus:** Richten Sie Richtlinien für den Datenlebenszyklus ein, um Daten je nach Alter und Nutzungsmustern automatisch zu verschieben oder zu löschen. Dies hilft bei der Verwaltung der Speicherkosten und sorgt dafür, dass der Datenspeicher seinem Wert entspricht.

## Entscheiden Sie sich AWS für Analysedienste

Nachdem Sie nun die Kriterien für die Bewertung Ihrer Analyseanforderungen kennen, können Sie entscheiden, welche AWS Analysedienste für Ihre Unternehmensanforderungen am besten geeignet sind. In der folgenden Tabelle sind die einzelnen Services nach gemeinsamen Funktionen und Geschäftszielen geordnet.

Kategorien	Wofür ist es optimiert?	Dienstleistungen
Vereinheitlichte Analytik und KI	Analytik und KI-Entwicklung  Optimiert für die Verwendung einer einzigen Entwicklungsumgebung, Amazon SageMaker Unified Studio,	<a href="#">Amazon SageMaker</a>

Kategorien	Wofür ist es optimiert?	Dienstleistungen
	für den Zugriff auf Daten, Analysen und KI-Funktionen.	
Datenverarbeitung	<p>Interaktive Analytik</p> <p>Optimiert für die Durchführung von Datenanalysen und -erkundungen in Echtzeit, sodass Benutzer Daten interaktiv abfragen und visualisieren können.</p>	<a href="#">Amazon Athena</a>
	<p>Verarbeitung großer Datenmengen</p> <p>Optimiert für die Verarbeitung, Übertragung und Transformation großer Datenmengen.</p>	<a href="#">Amazon EMR</a>
	<p>Datenkatalog</p> <p>Optimiert für die Bereitstellung detaillierter Informationen über die verfügbaren Daten, ihre Struktur, Eigenschaften und Beziehungen.</p>	<a href="#">AWS Glue</a>
	<p>Orchestrierung von Arbeitsabläufen</p> <p>Optimiert für die Erstellung, Planung und Überwachung von Datenworkflows mit Apache Airflow zur Koordination von Analyseprozessen und ETL-Jobs.</p>	<a href="#">Amazon MWAA</a>

Kategorien	Wofür ist es optimiert?	Dienstleistungen
Datenstreaming	<p>Verarbeitung von Streaming-Daten durch Apache Kafka</p> <p>Optimiert für die Verwendung von Apache Kafka-Datenebenenoperationen und die Ausführung von Open-Source-Versionen von Apache Kafka.</p>	<a href="#">Amazon MSK</a>
	<p>Verarbeitung in Echtzeit</p> <p>Optimiert für die schnelle und kontinuierliche Datenaufnahme und -aggregation, einschließlich Protokoll Daten der IT-Infrastruktur, Anwendungsprotokolle, soziale Medien, Marktdaten-Feeds und Web-Clickstream-Daten.</p>	<a href="#">Amazon Kinesis Data Streams</a>
	<p>Bereitstellung von Streaming-Daten in Echtzeit</p> <p>Optimiert für die Bereitstellung von Echtzeit-Streaming-Daten an Ziele wie Amazon S3, Amazon Redshift, OpenSearch Service, Splunk, Apache Iceberg Tables und alle benutzerdefinierten HTTP-Endpunkte oder HTTP-Endpunkte unterstützter Drittanbieter.</p>	<a href="#">Amazon Data Firehose</a>

Kategorien	Wofür ist es optimiert?	Dienstleistungen
	<p>Erstellung von Apache Flink-Anwendungen</p> <p>Optimiert für die Verwendung von Java, Scala, Python oder SQL zur Verarbeitung und Analyse von Streaming-Daten.</p>	<a href="#">Amazon Managed Service für Apache Flink</a>
Geschäftliche Intelligenz	<p>Dashboards und Visualisierungen</p> <p>Optimiert für die visuelle Darstellung komplexer Datensätze und die Bereitstellung von Abfragen Ihrer Daten in natürlicher Sprache.</p>	<a href="#">Quick Suite</a>
Analytik durchsuchen	<p>Verwaltete OpenSearch Cluster</p> <p>Optimiert für Protokollanalysen, Anwendungüberwachung in Echtzeit und Clickstream-Analyse.</p>	<a href="#">OpenSearch Amazon-Dienst</a>
Datenverwaltung	<p>Verwaltung des Datenzugriffs</p> <p>Optimiert für die Einrichtung der richtigen Verwaltung, Verfügbarkeit, Benutzerfreundlichkeit, Integrität und Sicherheit von Daten während ihres gesamten Lebenszyklus.</p>	<a href="#">Amazon DataZone</a>

Kategorien	Wofür ist es optimiert?	Dienstleistungen
Zusammenarbeit bei Daten	<p>Saubere Räume für sichere Daten</p> <p>Optimiert für die Zusammenarbeit mit anderen Unternehmen, ohne die zugrunde liegenden Rohdaten gemeinsam zu nutzen.</p>	<a href="#">AWS Clean Rooms</a>
Data Lake und Warehouse	<p>Einheitlicher Zugriff auf Data Lakes und Data Warehouses</p> <p>Basiert auf einer Lakehouse-Architektur zur Optimierung der Vereinheitlichung des Datenzugriffs über Amazon S3 S3-Data Lakes, Amazon Redshift Redshift-Data Warehouses, Betriebsdatenbanken sowie Drittanbieter- und Verbunddatenquellen.</p>	<a href="#">Amazon SageMaker</a>
	<p>Objektspeicher für Data Lakes</p> <p>Optimiert für die Bereitstellung einer Data-Lake-Grundlage mit praktisch unbegrenzter Skalierbarkeit und hoher Haltbarkeit.</p>	<a href="#">Amazon S3</a>

Kategorien	Wofür ist es optimiert?	Dienstleistungen
	<p>Data Warehousing</p> <p>Optimiert für das zentrale Speichern, Organisieren und Abrufen großer Mengen strukturierter und manchmal halbstrukturierter Daten aus verschiedenen Quellen innerhalb einer Organisation.</p>	<p><a href="#">Amazon Redshift</a></p>

## Verwenden Sie Analysedienste AWS

Sie sollten nun ein klares Verständnis Ihrer Geschäftsziele sowie des Volumens und der Geschwindigkeit der Daten haben, die Sie aufnehmen und analysieren werden, um mit dem Aufbau Ihrer Daten-Pipelines zu beginnen.

Um zu erfahren, wie Sie die einzelnen verfügbaren Dienste nutzen können, und mehr über sie erfahren, haben wir einen Weg bereitgestellt, um zu untersuchen, wie die einzelnen Dienste funktionieren. In den folgenden Abschnitten finden Sie Links zu ausführlicher Dokumentation, praktischen Tutorials und Ressourcen, die Ihnen den Einstieg von der grundlegenden Nutzung bis hin zu vertieften Einblicken in fortgeschrittenere Anwendungen erleichtern.

### Amazon Athena

- Erste Schritte mit Amazon Athena

Erfahren Sie, wie Sie Amazon Athena verwenden, um Daten abzufragen und eine Tabelle auf der Grundlage von in Amazon S3 gespeicherten Beispieldaten zu erstellen, die Tabelle abzufragen und die Ergebnisse der Abfrage zu überprüfen.

#### [Beginnen Sie mit dem Tutorial](#)

- Erste Schritte mit Apache Spark auf Athena

Verwenden Sie die vereinfachte Notebook-Oberfläche in der Athena-Konsole, um Apache Spark-Anwendungen mit Python oder Athena Notebook zu entwickeln. APIs

### [Fangen Sie mit dem Tutorial an](#)

- Katalogisieren und verwalten Sie Athena-Verbundabfragen mit der Amazon SageMaker Lakehouse-Architektur

Erfahren Sie, wie Sie über das Data Lakehouse in Amazon eine Verbindung zu Verbundabfragen herstellen, diese verwalten und für Daten ausführen, die in Amazon Redshift, DynamoDB und Snowflake gespeichert sind. SageMaker

### [Lesen Sie den Blog](#)

- Analysieren von Daten in Amazon S3 mit Athena

Erfahren Sie, wie Sie Athena für Logs von Elastic Load Balancers verwenden können, die als Textdateien in einem vordefinierten Format generiert wurden. Wir zeigen Ihnen, wie Sie eine Tabelle erstellen, die Daten in einem von Athena verwendeten Format partitionieren, sie in Parquet konvertieren und die Abfrageleistung vergleichen.

### [Lesen Sie den Blogbeitrag](#)

## AWS Clean Rooms

- Einrichtung AWS Clean Rooms

Erfahre, wie du es AWS Clean Rooms in deinem AWS Konto einrichtest.

### [Lesen Sie den Leitfaden](#)

- Mit AWS Entity Resolution on können Sie Dateneinblicke in Datensätzen mit mehreren Parteien gewinnen, AWS Clean Rooms ohne die zugrunde liegenden Daten gemeinsam nutzen zu müssen

Erfahren Sie, wie Sie mithilfe von Vorbereitung und Abgleich den Datenabgleich mit Mitarbeitern verbessern können.

### [Lesen Sie den Blogbeitrag](#)

- Wie differenzierter Datenschutz hilft, Erkenntnisse zu gewinnen, ohne Daten auf individueller Ebene preiszugeben

Erfahren Sie, wie AWS Clean Rooms Differential Privacy die Anwendung von Differential Privacy vereinfacht und zum Schutz der Privatsphäre Ihrer Benutzer beiträgt.



[Lesen Sie den Blog](#)

## Amazon Data Firehose

- Tutorial: Einen Firehose-Stream von der Konsole aus erstellen

Erfahren Sie, wie Sie das AWS-Managementkonsole oder ein AWS SDK verwenden, um einen Firehose-Stream zu Ihrem ausgewählten Ziel zu erstellen.

[Lesen Sie den Leitfaden](#)

- Daten an einen Firehose-Stream senden

Erfahren Sie, wie Sie verschiedene Datenquellen verwenden, um Daten an Ihren Firehose-Stream zu senden.

[Lesen Sie den Leitfaden](#)

- Transformieren Sie Quelldaten in Firehose

Erfahren Sie, wie Sie Ihre Lambda-Funktion aufrufen, um eingehende Quelldaten zu transformieren und die transformierten Daten an Ziele weiterzuleiten.

[Lesen Sie den Leitfaden](#)

## Amazon DataZone

- Erste Schritte mit Amazon DataZone

Erfahren Sie, wie Sie die DataZone Amazon-Root-Domain erstellen, die Datenportal-URL abrufen und die grundlegenden DataZone Amazon-Workflows für Datenproduzenten und Datenkonsumenten durchgehen.

[Fangen Sie mit dem Tutorial an](#)

- Ankündigung der allgemeinen Verfügbarkeit von Data Lineage in der nächsten Generation von Amazon und Amazon SageMaker DataZone

Erfahren Sie, wie Amazon die automatische Abstammungserfassung DataZone einsetzt, um sich auf die automatische Erfassung und Zuordnung von Abstammungsinformationen aus AWS Glue Amazon Redshift zu konzentrieren.

[Lesen Sie den Blog](#)

## Amazon EMR

- Erste Schritte mit Amazon EMR

Erfahren Sie, wie Sie mit Spark einen Beispielcluster starten und wie Sie ein einfaches PySpark Skript ausführen, das in einem Amazon S3 S3-Bucket gespeichert ist.

[Beginnen Sie mit dem Tutorial](#)

- Erste Schritte mit Amazon EMR auf Amazon EKS

Wir zeigen Ihnen, wie Sie mit der Nutzung von Amazon EMR auf Amazon EKS beginnen können, indem Sie eine Spark-Anwendung auf einem virtuellen Cluster bereitstellen.

[Erkunden Sie den Leitfaden](#)

- Erste Schritte mit EMR Serverless

Erfahren Sie, wie Amazon EMR Serverless eine serverlose Laufzeitumgebung bietet, die den Betrieb von Analyseanwendungen vereinfacht, die die neuesten Open-Source-Frameworks verwenden.

[Fangen Sie mit dem Tutorial an](#)

## AWS Glue

- Erste Schritte mit AWS Glue DataBrew

Erfahren Sie, wie Sie Ihr erstes DataBrew Projekt erstellen. Sie laden einen Beispieldatensatz, führen Transformationen für diesen Datensatz aus, erstellen ein Rezept für die Erfassung dieser Transformationen und führen einen Job aus, um die transformierten Daten in Amazon S3 zu schreiben.

[Beginnen Sie mit dem Tutorial](#)

- Transformieren Sie Daten mit AWS Glue DataBrew

Erfahren Sie mehr über AWS Glue DataBrew ein Tool zur visuellen Datenaufbereitung, das es Datenanalysten und Datenwissenschaftlern leicht macht, Daten zu bereinigen und zu

normalisieren, um sie für Analysen und maschinelles Lernen vorzubereiten. Erfahren Sie, wie Sie einen ETL-Prozess mithilfe von erstellen AWS Glue DataBrew.

### [Fangen Sie mit dem Labor an](#)

- AWS Glue DataBrew Tag des Eintauchens

Erfahren Sie, wie Sie AWS Glue DataBrew Daten für Analysen und maschinelles Lernen bereinigen und normalisieren können.

### [Fangen Sie mit dem Workshop an](#)

- Erste Schritte mit dem AWS Glue Data Catalog

Erfahren Sie, wie Sie Ihren ersten erstellen AWS Glue Data Catalog, der einen Amazon S3 S3-Bucket als Datenquelle verwendet.

### [Fangen Sie mit dem Tutorial an](#)

- Datenkatalog und Crawler in AWS Glue

Erfahren Sie, wie Sie die Informationen im Datenkatalog verwenden können, um Ihre ETL-Jobs zu erstellen und zu überwachen.

### [Erkunden Sie den Leitfaden](#)

## Amazon Kinesis Data Streams

- Tutorials „Erste Schritte“ für Amazon Kinesis Data Streams

Erfahren Sie, wie Sie Aktien Daten in Echtzeit verarbeiten und analysieren.

### [Fangen Sie mit den Tutorials an](#)

- Architekturmuster für Echtzeitanalysen mit Amazon Kinesis Data Streams, Teil 1

Erfahren Sie mehr über gängige Architekturmuster in zwei Anwendungsfällen: Zeitreihendatenanalyse und ereignisgesteuerte Microservices.

### [Lesen Sie den Blog](#)

- Architekturmuster für Echtzeitanalysen mit Amazon Kinesis Data Streams, Teil 2

Erfahren Sie mehr über KI-Anwendungen mit Kinesis Data Streams in drei Szenarien: Generative Business Intelligence in Echtzeit, Empfehlungssysteme in Echtzeit sowie Datenstreaming und Inferenz für das Internet der Dinge.

[Lesen Sie den Blog](#)

## Amazon Managed Service for Apache Flink

- Was ist Amazon Managed Service für Apache Flink?

Verstehen Sie die grundlegenden Konzepte von Amazon Managed Service für Apache Flink.

[Erkunden Sie den Leitfaden](#)

- Workshop zu Amazon Managed Service für Apache Flink

In diesem Workshop erfahren Sie, wie Sie eine Flink-Anwendung mit Amazon Managed Service für Apache Flink bereitstellen, betreiben und skalieren.

[Nehmen Sie am virtuellen Workshop teil](#)

## Amazon MSK

- Erste Schritte mit Amazon MSK

Erfahren Sie, wie Sie einen Amazon MSK-Cluster erstellen, Daten erzeugen und nutzen und den Zustand Ihres Clusters mithilfe von Metriken überwachen.

[Beginnen Sie mit dem Leitfaden](#)

- Amazon MSK-Werkstatt

Gehen Sie in diesem praxisnahen Amazon MSK-Workshop in die Tiefe.

[Beginnen Sie mit dem Workshop](#)

## Amazon MWAA

- Erste Schritte mit Amazon MWAA

Erfahren Sie, wie Sie Ihre erste MWAA-Umgebung erstellen, eine DAG auf Amazon S3 hochladen und Ihren ersten Workflow ausführen.

### [Beginnen Sie mit dem Tutorial](#)

- Aufbau von Daten-Pipelines mit Amazon MWAA

Erfahren Sie, wie Sie end-to-end Daten-Pipelines erstellen, die andere AWS Analysedienste wie Glue, EMR und Redshift orchestrieren. In diesem Blogbeitrag wird ein optimierter, konfigurationsgesteuerter Ansatz zur Orchestrierung von dbt Core-Jobs mithilfe von MWAA und Cosmos untersucht, wobei Jobs Transformationen auf Amazon Redshift ausführen.

### [Lesen Sie den Blogbeitrag](#)

- Amazon MWAA-Workshop

In praktischen Übungen erfahren Sie, wie Sie Amazon MWAA für die Orchestrierung von Daten-Workflows bereitstellen, konfigurieren und verwenden.

### [Beginnen Sie mit dem Workshop](#)

- Bewährte Methoden für Amazon MWAA

Lernen Sie Architekturmuster und bewährte Methoden für die Verwendung von Amazon MWAA in Ihren Analytics-Workflows kennen.

### [Lesen Sie den Leitfaden](#)

## OpenSearch Service

- Erste Schritte mit OpenSearch Service

Erfahren Sie, wie Sie Amazon OpenSearch Service verwenden, um eine Testdomain zu erstellen und zu konfigurieren.

### [Fangen Sie mit dem Tutorial an](#)

- Visualisieren von Kundendienstankufen mit OpenSearch Service und Dashboards OpenSearch

Hier finden Sie eine vollständige Beschreibung der folgenden Situation: Ein Unternehmen erhält eine bestimmte Anzahl von Kundendienstankufen und möchte diese analysieren. Wie lauteten die Themen der einzelnen Anrufe? Wie viele Gespräche waren positiv? Wie viele Gespräche

waren negativ? Wie können Vorgesetzte nach den Transkripten der Anrufe suchen oder diese überprüfen?

### [Fangen Sie mit dem Tutorial an](#)

- Workshop „Erste Schritte mit Amazon OpenSearch Serverless“

Erfahren Sie, wie Sie eine neue Amazon OpenSearch Serverless-Domain in der AWS Konsole einrichten. Erkunden Sie die verschiedenen verfügbaren Arten von Suchanfragen, entwerfen Sie ansprechende Visualisierungen und erfahren Sie, wie Sie Ihre Domain und Dokumente auf der Grundlage der zugewiesenen Benutzerrechte schützen können.

### [Fangen Sie mit dem Workshop an](#)

- Kostenoptimierte Vektordatenbank: Einführung in die Quantisierungstechniken von Amazon OpenSearch Service

Erfahren Sie, wie OpenSearch Service Skalar- und Produktquantisierungstechniken unterstützt, um die Speichernutzung zu optimieren und die Betriebskosten zu senken.

### [Lesen Sie den Blogbeitrag](#)

## Quick Suite

- Erste Schritte mit der Quick Suite-Datenanalyse

Erfahren Sie, wie Sie Ihre erste Analyse erstellen. Verwenden Sie Beispieldaten, um entweder eine einfache oder eine komplexere Analyse zu erstellen. Oder Sie können eine Verbindung zu Ihren eigenen Daten herstellen, um eine Analyse zu erstellen.

### [Erkunden Sie den Leitfaden](#)

- Visualisieren mit Quick Suite

Entdecken Sie die technische Seite von Business Intelligence (BI) und Datenvisualisierung mit AWS. Erfahren Sie, wie Sie Dashboards in Anwendungen und Websites einbetten und Zugriff und Berechtigungen sicher verwalten können.

### [Fangen Sie mit dem Kurs an](#)

- Quick Suite-Workshops

Verschaffen Sie sich mit Workshops einen Vorsprung auf Ihrer Quick Suite-Reise

## [Fangen Sie mit den Workshops an](#)

### Amazon Redshift

- Erste Schritte mit Amazon Redshift Serverless

Machen Sie sich mit dem grundlegenden Ablauf von Amazon Redshift Serverless vertraut, um serverlose Ressourcen zu erstellen, eine Verbindung zu Amazon Redshift Serverless herzustellen, Beispieldaten zu laden und dann Abfragen für die Daten auszuführen.

## [Erkunden Sie den Leitfaden](#)

- Vertiefender Workshop zu Amazon Redshift

Erkunden Sie eine Reihe von Übungen, die Benutzern den Einstieg in die Nutzung der Amazon Redshift Redshift-Plattform erleichtern.

## [Fangen Sie mit dem Workshop an](#)

### Amazon S3

- Erste Schritte mit Amazon S3

Erfahren Sie, wie Sie Ihr erstes DataBrew Projekt erstellen. Sie laden einen Beispieldatensatz, führen Transformationen für diesen Datensatz aus, erstellen ein Rezept für die Erfassung dieser Transformationen und führen einen Job aus, um die transformierten Daten in Amazon S3 zu schreiben.

## [Beginnen Sie mit dem Leitfaden](#)

### Amazon SageMaker

- Erste Schritte mit SageMaker

Erfahren Sie, wie Sie ein Projekt erstellen, Mitglieder hinzufügen und anhand des JupyterLab Beispielnoteizbuchs mit der Erstellung beginnen.

## [Lesen Sie den Leitfaden](#)

- Wir stellen die nächste Generation von Amazon vor SageMaker: Das Zentrum für all Ihre Daten, Analysen und KI

Erfahren Sie, wie Sie mit Datenverarbeitung, Modellentwicklung und generativer KI-App-Entwicklung beginnen können.

[Lesen Sie den Blog](#)

- Was ist SageMaker Unified Studio?

Erfahren Sie mehr über die Funktionen von SageMaker Unified Studio und wie Sie auf sie zugreifen können, wenn Sie Amazon verwenden SageMaker.

[Lesen Sie den Leitfaden](#)

- Erste Schritte mit der Lakehouse-Architektur von Amazon SageMaker

Erfahren Sie, wie Sie in Amazon ein Projekt erstellen und Daten für Ihre geschäftlichen Anwendungsfälle durchsuchen, hochladen und abfragen SageMaker.

[Lesen Sie den Leitfaden](#)

- Datenverbindungen in der Lakehouse-Architektur von Amazon SageMaker

Erfahren Sie, wie die Lakehouse-Architektur einen einheitlichen Ansatz für die Verwaltung von Datenverbindungen zwischen AWS Diensten und Unternehmensanwendungen bietet.

[Lesen Sie den Leitfaden](#)

- Katalogisieren und verwalten Sie föderierte Athena-Abfragen mit der SageMaker Lakehouse-Architektur

Erfahren Sie, wie Sie für Ihre Amazon-Projekte in Amazon Redshift, DynamoDB und Snowflake gespeicherte Verbundabfragen herstellen, diese verwalten und ausführen können. SageMaker

[Lesen Sie den Blog](#)



# Erkunden Sie Möglichkeiten zur Nutzung von AWS Analysediensten

## Editable architecture diagrams

### Referenzarchitekturdiagramme

Erkunden Sie Architekturdiagramme, mit denen Sie Ihre Analyselösungen entwickeln, skalieren und testen können AWS.

[Erkunden Sie die Referenzarchitekturen für Analysen](#)

## Ready-to-use code

Ausgewählte Lösung	AWS Lösungen
Skalierbare Analytik mit Apache Druid auf AWS	Entdecken Sie vorkonfigurierte, einsatzbereite Lösungen und deren Implementierungsleitfäden, entwickelt von AWS
Stellen Sie AWS integrierten Code bereit, der Sie bei der Einrichtung, dem Betrieb und der Verwaltung von Apache Druid auf einer kostengünstigen AWS, hochverfügbaren, belastbaren und fehlertoleranten Hosting-Umgebung unterstützt.	<a href="#">Erkunden Sie alle AWS Sicherheits-, Identitäts- und Governance-Lösungen</a>
<a href="#">Erkunden Sie diese Lösung</a>	

## Documentation

Whitepapers zur Analytik	AWS Blog über große Datenmengen
In Whitepapers finden Sie weitere Einblicke und bewährte Verfahren zur Auswahl, Implementierung und Nutzung der Analysedienste	Entdecken Sie Blogbeiträge, die sich mit bestimmten Big-Data-Anwendungsfällen befassen.

enste, die am besten zu Ihrem Unternehmen passen.

[Entdecken Sie Whitepapers zu Analysen](#)

[Erkunden Sie den AWS Big Data-Blog](#)

# Dokumentverlauf

In der folgenden Tabelle werden die wichtigen Änderungen an diesem Entscheidungsleitfaden beschrieben. Für Benachrichtigungen über Aktualisierungen dieses Handbuchs können Sie einen RSS-Feed abonnieren.

Änderung	Beschreibung	Datum
<a href="#">re:Invent-Updates</a>	Die Links im Entscheidungsleitfaden wurden aktualisiert und Amazon Managed Workflows für Apache Airflow hinzugefügt.	24. September 2025
<a href="#">re:Invent-Aktualisierungen</a>	Die Verweise auf Amazon SageMaker, Amazon SageMaker Unified Studio (nicht mehr als Vorschauversion) und Amazon SageMaker Lakehouse wurden im gesamten Entscheidungsleitfaden aktualisiert.	9. September 2025
<a href="#">re:Invent-Aktualisierungen</a>	SageMaker AI Unified Studio hinzugefügt und. AWS Clean Rooms Durchgehend aktualisiertes Dokument mit neuen KI-Funktionen und -Fähigkeiten.	20. Februar 2025
<a href="#">Erste Veröffentlichung</a>	Leitfaden zuerst veröffentlicht.	17. November 2023

Die vorliegende Übersetzung wurde maschinell erstellt. Im Falle eines Konflikts oder eines Widerspruchs zwischen dieser übersetzten Fassung und der englischen Fassung (einschließlich infolge von Verzögerungen bei der Übersetzung) ist die englische Fassung maßgeblich.