



Definition von Amazon S3 S3-Bucket- und Pfadnamen für Data-Lake-Ebenen

AWS Prescriptive Guidance



AWS Prescriptive Guidance: Definition von Amazon S3 S3-Bucket- und Pfadnamen für Data-Lake-Ebenen

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Die Handelsmarken und die 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

Einführung	1
Zielgruppe	2
Gezielte Geschäftsergebnisse	2
Empfohlene Datenschichten	3
Benennen Amazon S3 S3-Buckets	5
Landezone	6
Rohschicht	7
Stufe: Ebene	8
Analyse-Ebene	10
Zuordnung von Buckets zu IAM-Richtlinien	12
Umgang mit sensiblen Daten	14
Verwendung einer landing zone zum Maskieren sensibler Daten	14
Häufig gestellte Fragen	17
Welchen Namen sollte ich für einen Bucket mit mehreren Regionen verwenden?	17
Muss ich Raw, Stage und Analytics als Namen für meine Data-Lake-Ebenen verwenden?	17
Ist es möglich, einen Amazon S3 S3-Bucket umzubenennen?	17
Was passiert, wenn ich einen Bucket lösche und den Namen wiederverwenden möchte?	17
Gibt es Einschränkungen in Bezug darauf, was ich in meinen Bucket- oder Pfadnamen aufnehmen kann?	18
Kann ich in meinem Data Lake mehr Ebenen als die landing zone-, Raw-, Stage- und Analytics-Ebenen verwenden?	18
Was passiert, wenn ich meine Parameter nicht definiert habe?	18
Wie kann ich die Kosten auf Ebene der Geschäftsbereiche verfolgen?	18
Welche Funktionen sollte ich bei der Erstellung eines Benennungsstandards berücksichtigen?	19
Ressourcen	20
Dokumentverlauf	21
.....	xxii

Definition von Amazon S3 S3-Bucket- und Pfadnamen für Data-Lake-Ebenen

Andres Cantor, Amazon Web Services

April 2025 (Geschichte der [Dokumente](#))

Dieses Handbuch hilft Ihnen dabei, einen konsistenten Benennungsstandard für Amazon Simple Storage Service (Amazon S3) -Buckets und -Pfade in Data Lakes zu erstellen, die AWS Cloud auf dem gehostet werden. Der Benennungsstandard des Leitfadens für Amazon S3 S3-Buckets und -Pfade hilft Ihnen dabei, die Steuerung und Observability in Ihren Data Lakes zu verbessern, die Kosten nach Datenschichten zu identifizieren und AWS-Konto bietet einen Ansatz für die Benennung AWS Identity and Access Management (IAM) von Rollen und Richtlinien.

Wir empfehlen, dass Sie mindestens drei Datenschichten in Ihren Data Lakes verwenden und dass jede Ebene einen separaten Amazon S3 S3-Bucket verwendet. In einigen Anwendungsfällen sind jedoch möglicherweise ein zusätzlicher Amazon S3 S3-Bucket und eine zusätzliche Datenschicht erforderlich, je nachdem, welche Datentypen Sie generieren und speichern. Wenn Sie beispielsweise sensible Daten speichern, empfehlen wir Ihnen, eine Landingzone-Datenschicht und einen separaten Amazon S3 S3-Bucket zu verwenden. In der folgenden Liste werden die drei empfohlenen Datenschichten für Ihren Data Lake beschrieben:

- Rohdatenschicht — Enthält Rohdaten und ist die Ebene, in der Daten ursprünglich aufgenommen werden. Wenn möglich, empfehlen wir, das ursprüngliche Dateiformat beizubehalten und die Versionierung im Amazon S3 S3-Bucket zu aktivieren.
- Stage-Datenschicht — Enthält verarbeitete Zwischendaten, die für den Verbrauch optimiert sind (z. B. von CSV in Apache Parquet konvertierte Rohdateien oder Datentransformationen). Ein AWS Glue Job liest die Dateien aus der Rohschicht und validiert die Daten. Der AWS Glue Job speichert dann die Daten in einer Datei im Apache Parquet-Format, und die Metadaten werden in einer Tabelle im gespeichert. AWS Glue Data Catalog
- Analytics-Datenschicht — Enthält die aggregierten Daten für Ihre spezifischen Anwendungsfälle in einem benutzerfreundlichen Format wie Apache Parquet.

Zielgruppe

Die Empfehlungen dieses Handbuchs basieren auf den Erfahrungen der Autoren bei der Implementierung von Data Lakes mit dem [Serverless Data Lake Framework \(SDLF\)](#) und richten sich an Datenarchitekten, Dateningenieure oder Lösungsarchitekten, die einen Data Lake auf dem einrichten möchten. AWS Cloud Stellen Sie jedoch sicher, dass Sie den Ansatz dieses Handbuchs an die Richtlinien und Anforderungen Ihres Unternehmens anpassen.

Der Leitfaden umfasst die folgenden Abschnitte:

- [Empfohlene Datenschichten](#)
- [Benennen von Amazon S3 S3-Buckets in Ihren Datenschichten](#)
- [Zuordnen von Amazon S3 S3-Buckets zu IAM-Richtlinien in Ihrem Data Lake](#)
- [Umgang mit sensiblen Daten](#)

Gezielte Geschäftsergebnisse

Nach der Implementierung eines Benennungsstandards für Amazon S3 S3-Buckets und -Pfade in Data Lakes auf dem sollten Sie mit den AWS Cloud folgenden Ergebnissen rechnen:

- Verbesserte Steuerung in Ihrem Data Lake durch die Möglichkeit, differenzierte Zugriffsrichtlinien für die Buckets bereitzustellen
- Bessere Transparenz Ihrer Gesamtkosten für Einzelpersonen AWS-Konten durch Verwendung der entsprechenden AWS-Konto ID im Amazon S3 S3-Bucket-Namen und für Datenschichten durch Verwendung von [Kostenzuweisungs-Tags](#) für die Buckets
- Kostengünstigere Datenspeicherung durch die Verwendung von schichtbasierter Versionierung und pfadbasierten Lebenszyklusrichtlinien
- Erfüllen Sie die Sicherheitsanforderungen für Datenmaskierung und Datenverschlüsselung
- Vereinfachen Sie die Datenquellenverfolgung, indem Sie Entwicklern mehr Einblick in den AWS-Region und AWS-Konto den zugrunde liegenden Datenspeicher geben

Empfohlene Datenschichten

Wenn Sie mit nicht sensiblen Daten arbeiten, z. B. mit Daten, die keine personenbezogenen Daten (PII) enthalten, empfehlen wir Ihnen, mindestens drei verschiedene Datenschichten in einem Data Lake auf dem zu verwenden. AWS Cloud

Je nach Komplexität der Daten und Anwendungsfällen benötigen Sie jedoch möglicherweise zusätzliche Ebenen. Wenn Sie beispielsweise mit sensiblen Daten wie PII-Daten arbeiten, empfehlen wir Ihnen, einen zusätzlichen Amazon Simple Storage Service (Amazon S3) -Bucket als landing zone zu verwenden. Anschließend maskieren Sie die Daten, bevor sie in die Rohdatenebene verschoben werden. Weitere Informationen dazu finden Sie im Abschnitt [Umgang mit sensiblen Daten](#) in diesem Handbuch.

Jede Datenschicht muss über einen individuellen Amazon S3 S3-Bucket verfügen. In der folgenden Tabelle werden die empfohlenen Datenschichten beschrieben.

Name der Datenschicht	Beschreibung	Beispiel einer Strategie für eine Lebenszyklusrichtlinie
Roh	Enthält die unverarbeiteten Rohdaten. Daten werden in den Data Lake in dieser Schicht aufgenommen. Wenn möglich, sollten Sie das ursprüngliche Dateiformat beibehalten und die Versionierung im Amazon S3 S3-Bucket aktivieren.	Verschieben Sie Dateien nach einem Jahr in die Amazon S3 S3-Speicherklasse für seltenen Zugriff (IA). Archivieren Sie sie nach zwei Jahren in Amazon S3 IA in Amazon Glacier-Speicherklassen.
Stage	Enthält verarbeitete Zwischendaten, die für den Verbrauch optimiert sind (z. B. konvertierte Rohdateien von CSV in Apache Parquet oder Datentransformationen).	Daten können nach einem bestimmten Zeitraum oder gemäß den Anforderungen Ihres Unternehmens gelöscht werden.

Name der Datenschicht	Beschreibung	Beispiel einer Strategie für eine Lebenszyklusrichtlinie
	Ein AWS Glue Job liest die Dateien aus der Rohschicht und validiert die Daten. Der AWS Glue Job speichert dann die Daten in einer Datei im Apache Parquet-Format, und die Metadaten werden in einer Tabelle im gespeichert. AWS Glue Data Catalog	Einige Datenderivate, wie z. B. eine Apache Avro-Transformation eines ursprünglichen JSON-Formats, können nach kürzerer Zeit, z. B. nach 90 Tagen, aus dem Data Lake entfernt werden.
Analysen	Enthält die aggregierten Daten für Ihre spezifischen Anwendungsfälle in einem benutzerfreundlichen Format, z. B. Apache Parquet.	Daten können nach Amazon S3 IA verschoben und dann nach einem bestimmten Zeitraum oder gemäß den Anforderungen Ihres Unternehmens gelöscht werden.

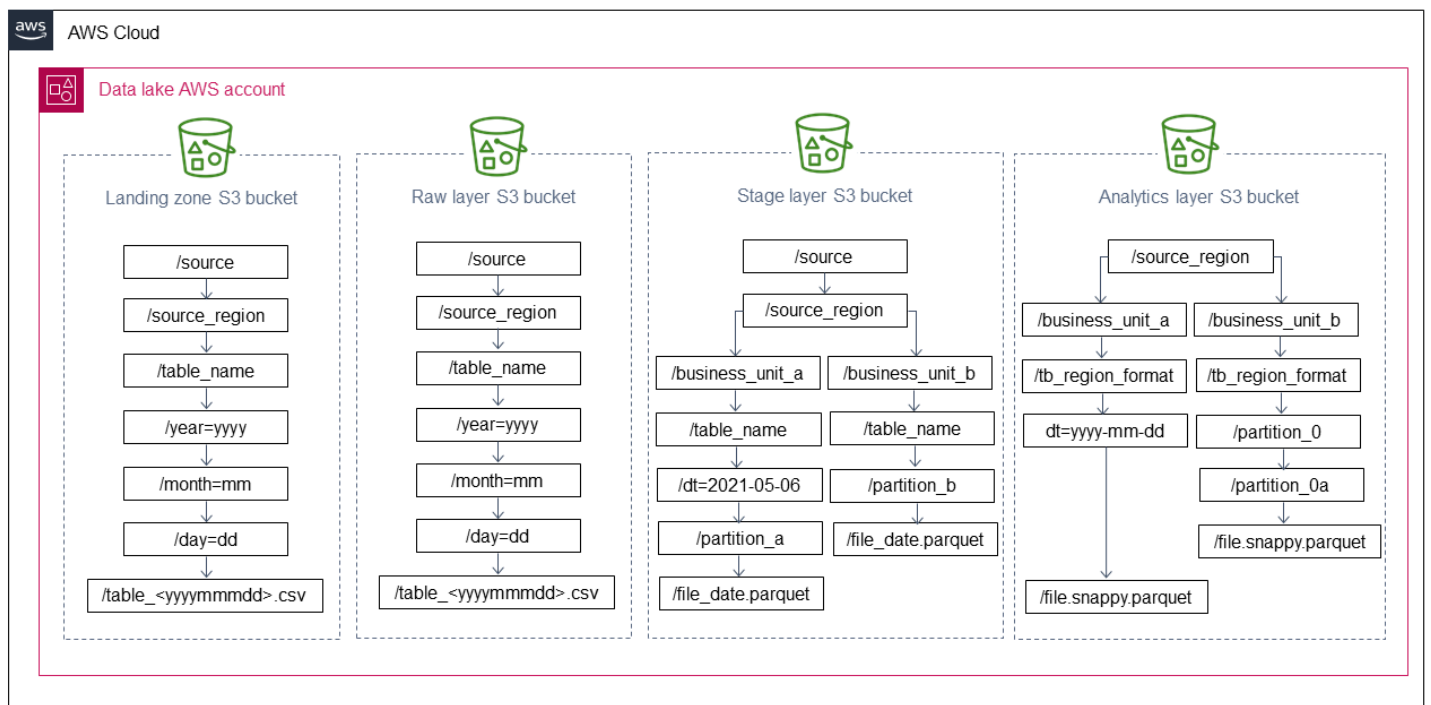
 Note

Sie müssen alle empfohlenen Strategien für die Lebenszykluspolitik anhand Ihrer organisatorischen Anforderungen, regulatorischen Anforderungen, Abfragemuster und Kostenüberlegungen bewerten.

Benennen von Amazon S3 S3-Buckets in Ihren Datenschichten

In den folgenden Abschnitten finden Sie Benennungsstrukturen für Amazon Simple Storage Service (Amazon S3) -Buckets in Ihren Data Lake-Ebenen. Sie können die Amazon S3 S3-Bucket- und Pfadnamen jedoch an die Anforderungen Ihrer Organisation anpassen. Wir empfehlen, separate Buckets für jede einzelne Ebene zu erstellen, da die Anforderungen an Archivierung, Versionierung, Zugriff und Verschlüsselung für jede Ebene variieren können.

Das folgende Diagramm zeigt die empfohlene Benennungsstruktur für Amazon S3 S3-Buckets in den empfohlenen Data-Lake-Layern. Die Benennungsstruktur trennt mehrere Geschäftsbereiche, Dateiformate und Partitionen.



⚠ Important

Amazon S3 S3-Buckets müssen den Benennungsrichtlinien der [Bucket-Benennungsregeln](#) in der Amazon S3 S3-Dokumentation entsprechen.

Sie können Datenpartitionen an die Anforderungen Ihres Unternehmens anpassen. Sie sollten jedoch Paare aus Kleinbuchstaben und Schlüssel/Wert verwenden (z. B. `year=yyyy` anstelle von `yyyy`), damit Sie den Katalog mit dem Befehl aktualisieren können. `MSCK REPAIR TABLE`

Die Definition einer Partitionsstrategie hängt von der Art Ihrer Daten und vor allem von der Art Ihrer Benutzerabfragen ab. Wir empfehlen Ihnen, die Verbrauchs- und Datenverarbeitungsmuster zu analysieren, um die für Ihr Unternehmen am besten geeignete Strategie zu finden. Im Allgemeinen ist es sinnvoll, höhere Hierarchieebenen, z. B. `year=yyyy`, und `month=mmday=dd`, auf der Rohdatenebene und niedrigere Hierarchieebenen auf Verbrauchsdatenebenen wie der Stufe- und Analyseebene vorzusehen. Dies liegt daran, dass Rohdatenschichten in der Regel nicht die komplexen Nutzungsmuster von Datenverarbeitungspipelines aufweisen.

Landezone Amazon S3 S3-Bucket

Sie benötigen einen Amazon S3 S3-Bucket für Ihre landing zone, wenn sensible Datensätze Elemente enthalten, die maskiert werden müssen, bevor Daten in den Raw-Bucket verschoben werden.

Die folgende Tabelle enthält die Benennungsstruktur, eine Beschreibung der Benennungsstruktur und ein Namensbeispiel für den Amazon S3 S3-Bucket in Ihrem landing zone Zone-Layer.

Benennungsformat	Beispiel
<code>s3://companyname-landingzoneawsregion-awsaccount uniqueid/source/source_region/table/year=yyyy/month=mm/day=dd/table_<yearmonthday>.avro csv</code>	<code>s3://anycompany-landingzoneuseast1-12345-dev/socialmedia/us/tb_products/year=2021/month=03/day=01/products_20210301.csv</code>
<code>companyname</code> — Der Name der Organisation (optional) <code>awsregion</code> — Der AWS-Region, wie <code>us-east-1</code> oder <code>sa-east-1</code> <code>awsaccount uniqueid</code> — Die eindeutige Kennung oder AWS-Konto ID	

Benennungsformat	Beispiel
• <code>env</code>— Die Bereitstellungsumgebung, wie <code>dev</code>, <code>test</code>, oder <code>prod</code> • <code>source</code>— Die Quelle oder der Inhalt, z. B. MySQL-Datenbank, E-Commerce oder SAP • <code>source_region</code> — Globale Geschäftsregion, wie <code>us</code> oder <code>asia</code> • <code>table</code> – <code>tb_customer</code> , <code>tb_transactions</code> oder <code>tb_products</code>	

Amazon S3 S3-Bucket mit Rohschicht

Die Rohdatenschicht enthält aufgenommene Daten, die nicht transformiert wurden und in ihrem ursprünglichen Dateiformat wie JSON oder CSV vorliegen. Diese Daten sind in der Regel nach Datenquelle und Datum geordnet, an dem sie in den Amazon S3 S3-Bucket der Rohdatenschicht aufgenommen wurden.

Die folgende Tabelle enthält die Benennungsstruktur, eine Beschreibung der Benennungsstruktur und ein Namensbeispiel für den Amazon S3 S3-Bucket in Ihrer Rohdatenschicht.

Benennungsformat	Beispiel
<code>s3://companyname-raw-awsregion-awsaccount uniqid-env/source/source_region/table/year=yyyy/month=mm/day=dd/table_<yearmonthday>.avro csv</code> • <code>companyname</code> — Der Name der Organisation (optional)	<code>s3://anycompany-raw-useast1-12345-dev/socialmedia/us/tb_products/year=2021/month=03/day=01/products_20210301.csv</code>

Benennungsformat	Beispiel
• <code>awsregion</code> — Der AWS-Region, wie <code>us-east-1</code> oder <code>sa-east-1</code> • <code>awsaccount uniqid</code> — Die eindeutige Kennung oder AWS-Konto ID • <code>env</code>— Die Bereitstellungsumgebung, wie <code>dev</code>, <code>test</code>, oder <code>prod</code> • <code>source</code>— Die Quelle oder der Inhalt, z. B. MySQL-Datenbank, E-Commerce oder SAP • <code>source_region</code> — Globale Geschäftsregion, wie <code>us</code> oder <code>asia</code> • <code>table</code> – <code>tb_customer</code> , <code>tb_transactions</code> oder <code>tb_products</code>	

Amazon S3 S3-Bucket auf Stufenebene

Daten in der Stageebene werden aus der Rohschicht gelesen und transformiert (z. B. mithilfe eines AWS Glue oder eines Amazon EMR-Jobs). Bei diesem Prozess werden die Daten validiert (z. B. durch Überprüfung von Datentypen und Headern) und anschließend in einem nutzbaren Dateiformat wie Apache Parquet gespeichert. Die Metadaten werden in einer Tabelle im gespeichert. [AWS Glue Data Catalog](#)

Die folgende Tabelle enthält die Benennungsstruktur, eine Beschreibung der Benennungsstruktur und ein Namensbeispiel für den Amazon S3 S3-Bucket in Ihrer Stage-Datenschicht.

Benennungsformat	Beispiel
<code>s3://companyname-stageawsregion-awsaccount uniqidenv/source/</code>	<code>s3://anycompany-stagesaeast1-12345-dev/sap/br/customer</code>

Benennungsformat	Beispiel
source_region/ business_unit/ table/<partitions>/table_<table_name>_<yearmonthday>.snap • companyname — Der Name der Organisation (optional) • awsregion — Der AWS-Region, wie us-east-1 oder sa-east-1 • awsaccount unqid — Die eindeutige Kennung oder AWS-Konto ID • env— Die Bereitstellungsumgebung, wie dev, test, oder prod • source— Die Quelle oder der Inhalt, z. B. MySQL-Datenbank, E-Commerce oder SAP • source_region — Globale Geschäftsregion, wie us oder asia • business_unit — Die Geschäftseinheit, für die die Daten verarbeitet werden • table – tb_customer , tb_transactions oder tb_products • partitions — Partitionen, die dem Verbraucher die beste Leistung bieten und es der Abfrage-Engine ermöglichen, vollständige Datenscans zu vermeiden	s/validated/dt=2021-03-01/table_customers_20210301.snap ppy.parquet py.parquet

Analyseschicht Amazon S3 S3-Bucket

Die Analyseebene ähnelt der Stage-Ebene, da die Daten in einem verarbeiteten Dateiformat vorliegen, die Daten dann jedoch gemäß den Anforderungen Ihrer Organisation aggregiert werden.

Die folgende Tabelle enthält die Benennungsstruktur, eine Beschreibung der Benennungsstruktur und ein Namensbeispiel für den Amazon S3 S3-Bucket in Ihrer Analytics-Datenschicht.

Benennungsformat	Beispiel
<code>s3://companyname-analytics-awsregion-awsaccount uniqid-env/source_region/business_unit/tb_<region>_<table_name>_<file_format>/<partition_0>/<partition_1>/.../<partition_n>/xxxxx.<compression>.<file_format></code> • <code>companyname</code> — Der Name der Organisation (optional) • <code>awsregion</code> — Der AWS-Region, wie <code>us-east-1</code> oder <code>sa-east-1</code> • <code>awsaccount uniqid</code> — Die eindeutige Kennung oder AWS-Konto ID • <code>env</code>— Die Bereitstellungsumgebung, wie <code>dev</code>, <code>test</code>, oder <code>prod</code> • <code>source</code>— Die Quelle oder der Inhalt, z. B. MySQL-Datenbank, E-Commerce oder SAP	<code>s3://anycompany-analytics-useast1-12345-dev/us/sales/tb_us_customers_parquet/<partitions>/part-000001-20218c886790.c000.snappy.parquet</code>

Benennungsformat	Beispiel
• <code>source_region</code> — Globale Geschäftsregion, wie <code>us</code> oder <code>asia</code> • <code>business_unit</code> — Die Geschäftseinheit, für die die Daten verarbeitet werden • <code>table – tb_customer , tb_transactions</code> oder <code>tb_products</code> • <code>partitions</code> — Partitionen, die dem Verbraucher die beste Leistung bieten und es der Abfrage-Engine ermöglichen, vollständige Datenscans zu vermeiden	

Zuordnen von Amazon S3 S3-Buckets zu IAM-Richtlinien in Ihrem Data Lake

Wir empfehlen, dass Sie die Amazon Simple Storage Service (Amazon S3) -Buckets und -Pfade des Data Lakes AWS Identity and Access Management (IAM) -Richtlinien und -Rollen zuordnen. Sie verwenden die Bucket-Namen oder -Pfade in der IAM-Richtlinie oder im Rollennamen. Die folgende Tabelle zeigt ein Beispiel für einen Amazon S3 S3-Bucket-Namen und ein Beispiel für eine IAM-Richtlinie, die für den Zugriff auf diesen Bucket verwendet wird.

Beispiel für einen Objektpfad	Beispiel einer IAM-Richtlinie
Name des Amazon S3 S3-Buckets — <companyname>-raw-<aws_region>-<aws_accountid>-dev Amazon S3 S3-Bucket-Pfad — nosql/us/customers/year=2020/month=03/day=01/table_customers_20210301.csv	<pre>{ "Version" : "2012-10-17", "Statement" : [{ "Sid" : "s3-nosql-us-customers-get-list", "Effect" : "Allow", "Principal" : "*", "Action" : ["s3:GetObject", "s3:ListBucket"], "Resource" : ["arn:aws:s3:::<companyname>-raw-<aws_region>-<aws_accountid>-dev/*"] }] }</pre>

 Note

Dies ist ein Beispiel für eine IAM-Richtlinie, die den empfohlenen Benennungsstandard für Amazon S3 S3-Buckets zeigt. Sie sollten jedoch sicherstellen, dass Sie die Bucket-Richtlinien gemäß den Richtlinien und Anforderungen Ihrer Organisation korrekt konfigurieren.

Umgang mit sensiblen Daten

In der Regel enthalten sensible Daten personenbezogene Daten (PII) oder vertrauliche Informationen, die aus Compliance-Gründen oder aus rechtlichen Gründen geschützt werden müssen. Wenn die Verschlüsselung nur auf Zeilen- oder Spaltenebene erforderlich ist, empfehlen wir die Verwendung eines Landingzone-Layers. Dies sind teilweise sensible Daten.

Wenn jedoch der gesamte Datensatz als vertraulich eingestuft wird, empfehlen wir, separate Amazon Simple Storage Service (Amazon S3) -Buckets zu verwenden, um die Daten zu speichern. Dabei handelt es sich um hochsensible Daten. Diese separaten Amazon S3 S3-Buckets müssen für jede Datenschicht verwendet werden, und „sensibel“ sollte im Namen des Buckets enthalten sein.

Wir empfehlen, sensible Buckets mit AWS Key Management Service (AWS KMS) zu verschlüsseln, indem Sie die clientseitige Verschlüsselung verwenden. Sie müssen auch die clientseitige Verschlüsselung verwenden, um die Jobs zu verschlüsseln, die Ihre Daten transformieren. AWS Glue Client-side Die Verschlüsselung sollte für diese Buckets und die Rollen der Datenverarbeitungspipelines konfiguriert werden, z. B. für die IAM-Rolle für den Job. AWS Glue Diese Rollen müssen über die entsprechenden Berechtigungen verfügen, um den konfigurierten KMS-Schlüssel zu verwenden und in den Bucket zu lesen und in ihn zu schreiben.

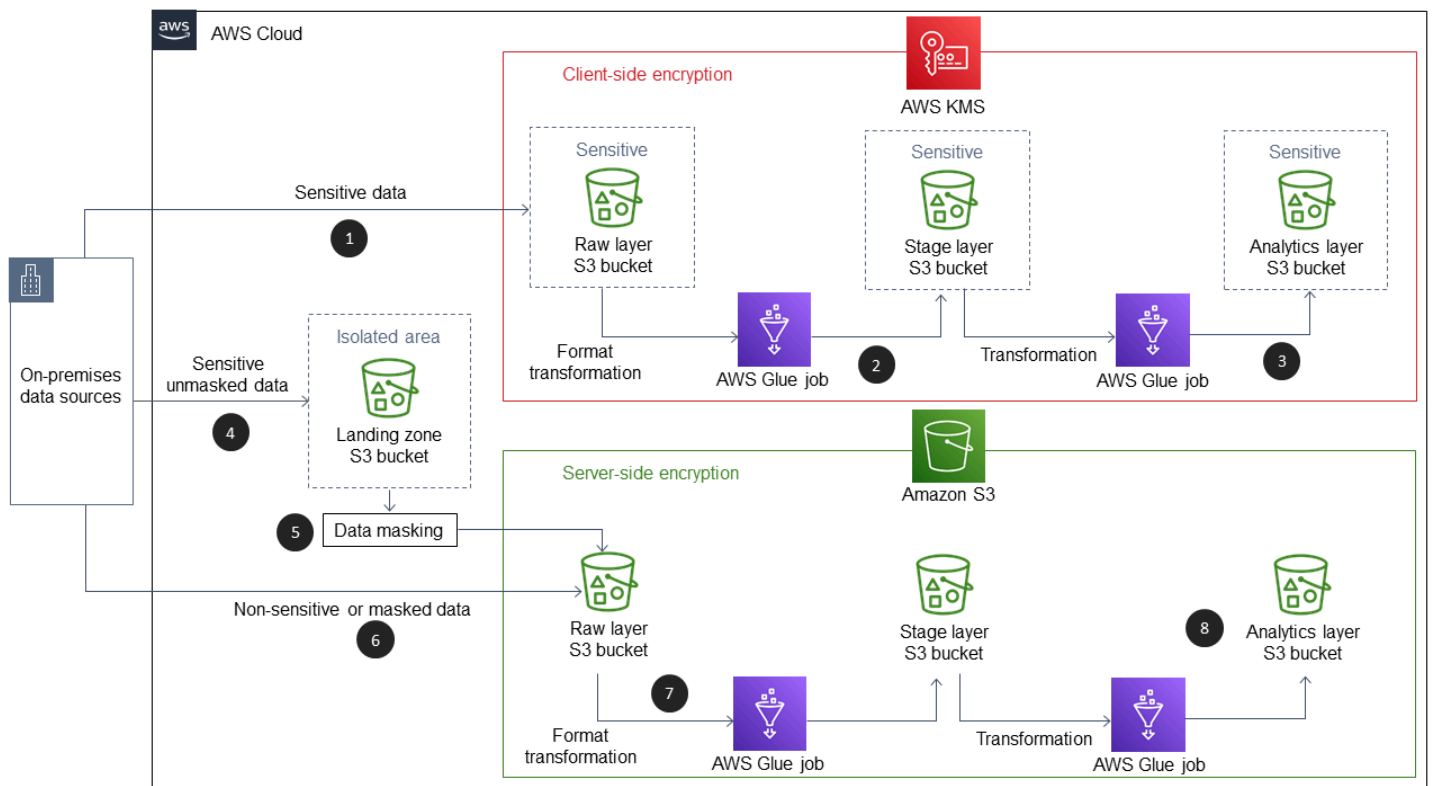
Verwendung einer landing zone zum Maskieren sensibler Daten

Sie können einen Landingzone-Layer für teilweise sensible Datasets verwenden (z. B. wenn eine Verschlüsselung nur auf Zeilen- oder Spaltenebene erforderlich ist). Diese Daten werden in den Amazon S3 S3-Bucket der Landing Zone aufgenommen und anschließend maskiert. Nachdem die Daten maskiert wurden, werden sie in den Amazon S3 S3-Bucket der Rohschicht aufgenommen. Dieser Bucket ist mit serverseitiger Verschlüsselung unter Verwendung von verwalteten Amazon S3 S3-Schlüsseln (SSE-S3) verschlüsselt. Bei Bedarf können Sie Daten auf Objektebene taggen.

Alle Daten, die bereits maskiert sind, können die landing zone umgehen und direkt in den Amazon S3 S3-Bucket der Rohschicht aufgenommen werden. Es gibt zwei Zugriffsebenen in der Phase- und Analyseebene für teilweise sensible Datensätze. Eine Ebene hat vollen Zugriff auf alle Daten und die andere Ebene hat nur Zugriff auf nicht sensible Zeilen und Spalten.

Das folgende Diagramm zeigt einen Data Lake, in dem teilweise sensible Datensätze eine landing zone verwenden, um die sensiblen Daten zu maskieren, hochsensible Datensätze jedoch separate,

verschlüsselte Amazon S3 S3-Buckets verwenden. Die landing zone wird mithilfe restriktiver IAM- und Bucket-Richtlinien isoliert, und die verschlüsselten Buckets verwenden die clientseitige Verschlüsselung mit AWS KMS



Das Diagramm zeigt den folgenden Workflow:

1. Highly-sensitive Daten werden an einen verschlüsselten Amazon S3 S3-Bucket in der Rohdatenschicht gesendet.
2. Ein AWS Glue Job validiert und transformiert die Daten in ein verbrauchsfertiges Format und platziert die Datei dann in einem verschlüsselten Amazon S3 S3-Bucket in der Stage-Ebene.
3. Ein AWS Glue Job aggregiert Daten gemäß den Geschäftsanforderungen und platziert die Daten in einem verschlüsselten Amazon S3 S3-Bucket in der Analyseebene.
4. Partially-sensitive Daten werden an den landing zone Bucket gesendet.
5. Vertrauliche Zeilen und Spalten werden maskiert, und die Daten werden dann in der Rohschicht an den Amazon S3 S3-Bucket gesendet.
6. Non-sensitive Daten werden in der Rohschicht direkt an den Amazon S3 S3-Bucket gesendet.
7. Ein AWS Glue Job validiert und transformiert die Daten in ein verbrauchsfertiges Format und platziert die Dateien im Amazon S3 S3-Bucket für die Stage-Ebene.

8. Ein AWS Glue Job aggregiert die Daten gemäß den Anforderungen Ihres Unternehmens und platziert die Daten in einem Amazon S3 S3-Bucket auf der Analyseebene.

Häufig gestellte Fragen

Dieser Abschnitt enthält Antworten auf häufig gestellte Fragen zur Definition von Amazon Simple Storage Service (Amazon S3) -Bucket- und Pfadnamen für Data-Lake-Ebenen auf dem AWS Cloud.

Welchen Namen sollte ich für einen Bucket mit mehreren Regionen verwenden?

Sie können unser empfohlenes Benennungsformat für Amazon S3 S3-Buckets verwenden und die AWS-Region ID ändern. Beispiele hierfür sind `examplecompany-raw-useast1-12345-dev` und `examplecompany-raw-uswest1-12345-dev`.

Muss ich Raw, Stage und Analytics als Namen für meine Data-Lake-Ebenen verwenden?

Nein, Sie können Ihre Ebenen nach Ihren Anforderungen benennen. Wir empfehlen jedoch dringend, einen Amazon S3 S3-Bucket für die Datenschicht zu verwenden, der die ursprünglichen Dateiformate enthält, und die Versionierung für diesen Bucket zu aktivieren.

Ist es möglich, einen Amazon S3 S3-Bucket umzubenennen?

Nein. Wenn Sie einen anderen Amazon S3 S3-Bucket-Namen verwenden möchten, müssen Sie einen neuen Bucket mit dem neuen Namen erstellen. Dies ist ein Grund, warum wir einen klar definierten und konsistenten Benennungsansatz für Amazon S3 S3-Buckets empfehlen.

Was passiert, wenn ich einen Bucket lösche und den Namen wiederverwenden möchte?

Wenn Sie einen Amazon S3 S3-Bucket löschen und einen neuen Bucket mit demselben Namen erstellen möchten, müssen Sie warten, bis der Name wieder verfügbar ist. Es kann 48 bis 72 Stunden dauern, bis die Änderungen wirksam werden. Es hat sich bewährt, mindestens 48 Stunden zu warten, bevor Sie einen neuen Bucket erstellen, der einen früheren Namen wiederverwendet. Amazon S3 S3-Bucket-Namen sind weltweit einzigartig und AWS-Konten haben alle denselben Namespace.

Gibt es Einschränkungen in Bezug darauf, was ich in meinen Bucket- oder Pfadnamen aufnehmen kann?

In Amazon S3 S3-Bucket-Namen sind nur Kleinbuchstaben, Zahlen, Bindestriche und Punkte zulässig. Bucket-Namen müssen 3—63 Zeichen lang sein, müssen mit einer Zahl oder einem Buchstaben beginnen und enden und dürfen kein IP-Adressformat haben. Die Namen müssen außerdem global eindeutig sein.

Für Amazon S3 S3-Bucket-Pfade können Sie Großbuchstaben verwenden, wir empfehlen jedoch, nur Kleinbuchstaben zu verwenden. Pfade können auch zusätzliche Symbole enthalten, wir empfehlen jedoch, nur Unterstriche, Bindestriche, Schrägstriche und Zahlen zu verwenden.

Kann ich in meinem Data Lake mehr Ebenen als die landing zone-, Raw-, Stage- und Analytics-Ebenen verwenden?

Ja, Sie können so viele Ebenen verwenden, wie Sie möchten. Wir empfehlen jedoch, eine landing zone und eine Rohschicht für Ihre Rohdaten, eine Zwischenschicht für formatierte Daten und eine Ebene für stark modellierte Daten zu haben.

Was passiert, wenn ich meine Parameter nicht definiert habe?

Bestimmte Parameter, wie z. B. Geschäftseinheiten, müssen nicht in den Amazon S3 S3-Bucket-Namen aufgenommen werden, können aber Teil des Pfads sein. Das bedeutet, dass sie nicht sofort ermittelt werden müssen, da Pfade hinzugefügt werden können, nachdem ein Amazon S3 S3-Bucket erstellt wurde.

Wie kann ich die Kosten auf Ebene der Geschäftsbereiche verfolgen?

Das hängt von Ihrer Kontostrategie ab. Wenn Ihre Geschäftsbereiche separate haben AWS-Konten, können Sie Amazon S3 S3-Buckets Kostenzuweisungs-Tags zuweisen, die die Bucket-Kosten für jede Geschäftseinheit widerspiegeln.

Wenn Ihre Kontostrategie die Geschäftsbereiche nicht in verschiedene unterteilt AWS-Konten, können Sie für jede Geschäftseinheit unterschiedliche Bereiche verwenden. Fügen Sie die

Geschäftseinheit dem Bucket-Namen hinzu, z. B. `exampleco-businessunit1-raw-useast1-12345-dev`. Dies bedeutet jedoch, dass Sie viele Amazon S3 S3-Buckets verwalten müssen.

Welche Funktionen sollte ich bei der Erstellung eines Benennungsstandards berücksichtigen?

Sie müssen sicherstellen, dass Ihre Amazon S3 S3-Bucket-Namen Funktionen verwenden, die nur auf Bucket-Ebene verfügbar sind. Beispielsweise sind Kostenschilder, Bucket-Verschlüsselung und Versionierung Funktionen, die nur für einen gesamten Amazon S3 S3-Bucket verfügbar sind. Das bedeutet, dass sie für alle Objekte und Pfade im Bucket gelten.

Die Versionierung von Objekten ist ebenfalls ein wichtiges Merkmal, das berücksichtigt werden muss. Sie sollten die Versionierung für die Amazon S3 S3-Buckets Ihres Raw-Layers aktivieren. Dadurch wird sichergestellt, dass Sie bei Änderungen an den Daten auf frühere Versionen zugreifen können. Eine Versionierung ist jedoch möglicherweise nicht für alle Ebenen in Ihrem Data Lake erforderlich, und die Beibehaltung mehrerer Versionen kann zu unnötigen Kosten führen.

Ressourcen

- [AWS Glue Entwicklerhandbuch](#) (AWS Glue Dokumentation)
- [AWS Glue Komponenten](#) (AWS Glue Dokumentation)
- [Benennungsregeln für Buckets](#) (Amazon S3 S3-Dokumentation)
- [Serverseitige Verschlüsselung mit AWS KMS keys \(SSE-KMS\) verwenden](#) (Amazon S3 S3-Dokumentation)
- [Verwenden von S3-Bucket-Tags für die Kostenzuweisung](#) (Amazon S3 S3-Dokumentation)
- [Benutzerdefinierte Tags für die Kostenzuweisung](#) (AWS Fakturierung und Kostenmanagement Dokumentation)

Dokumentverlauf

In der folgenden Tabelle werden wichtige Änderungen in diesem Leitfaden beschrieben. Um Benachrichtigungen über zukünftige Aktualisierungen zu erhalten, können Sie einen [RSS-Feed](#) abonnieren.

Änderung	Beschreibung	Datum
Aktualisieren	Wir haben weitere Informationen zur Definition einer Partitionsstrategie hinzugefügt und die häufig gestellten Fragen aktualisiert.	22. April 2025
Erste Veröffentlichung	—	18. November 2021

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.