

AWS 결정 가이드

AWS 분석 서비스 선택

AWS 분석 서비스 선택: AWS 결정 가이드

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

Amazon의 상표 및 트레이드 드레스는 Amazon 외 제품 또는 서비스와 함께, Amazon 브랜드 이미지를 떨어뜨리거나 고객에게 혼동을 일으킬 수 있는 방식으로 사용할 수 없습니다. Amazon이 소유하지 않은 기타 모든 상표는 Amazon과 제휴 관계이거나 관련이 있거나 후원 관계와 관계없이 해당 소유자의 자산입니다.

Table of Contents

결정 가이드	1
소개	1
이해	2
고려 사항	6
선택	13
사용	16
탐색	25
문서 기록	27
.....	xxviii

AWS 분석 서비스 선택

첫 번째 단계 수행

용도	조직에 가장 적합한 AWS 분석 서비스를 결정하는 데 도움이 됩니다.
최종 업데이트 날짜	2025년 9월 24일
적용 대상 서비스	<ul style="list-style-type: none">Amazon AthenaAWS Clean RoomsAmazon Data FirehoseAmazon DataZoneAmazon EMRAWS GlueAmazon Kinesis Data StreamsAmazon Managed Service for Apache FlinkAmazon Managed Streaming for Apache KafkaAmazon Managed Workflows for Apache AirflowAmazon OpenSearch ServiceQuick SuiteAmazon RedshiftAmazon S3Amazon SageMaker

소개

데이터는 현대 비즈니스의 기초입니다. 사람과 애플리케이션은 새롭고 다양한 소스에서 가져온 데이터에 안전하게 액세스하고 분석해야 합니다. 또한 데이터 양이 지속적으로 증가하여 조직이 필요한 모든 데이터를 캡처, 저장 및 분석하는 데 어려움을 겪을 수 있습니다.

이러한 문제를 해결하려면 타사 데이터를 포함한 분석 및 인사이트를 위해 모든 데이터 파일로를 세분화하고 조직 내 모든 사람이 엔드 end-to-end 거버넌스를 통해 한 곳에서 액세스할 수 있는 최신 데이터 아키텍처를 구축해야 합니다. 또한 분석 및 기계 학습(ML) 시스템을 연결하여 예측 분석을 활성화하는 것도 점점 더 중요합니다.

이 결정 가이드는 AWS 서비스에 최신 데이터 아키텍처를 구축하기 위한 올바른 질문을 하는 데 도움이 됩니다. 데이터 파일로(데이터 레이크 및 데이터 웨어하우스 연결), 시스템 파일로(ML 및 분석 연결) 및 인력 파일로(조직 내 모든 사람이 데이터를 사용할 수 있도록 함)를 세분화하는 방법을 설명합니다.

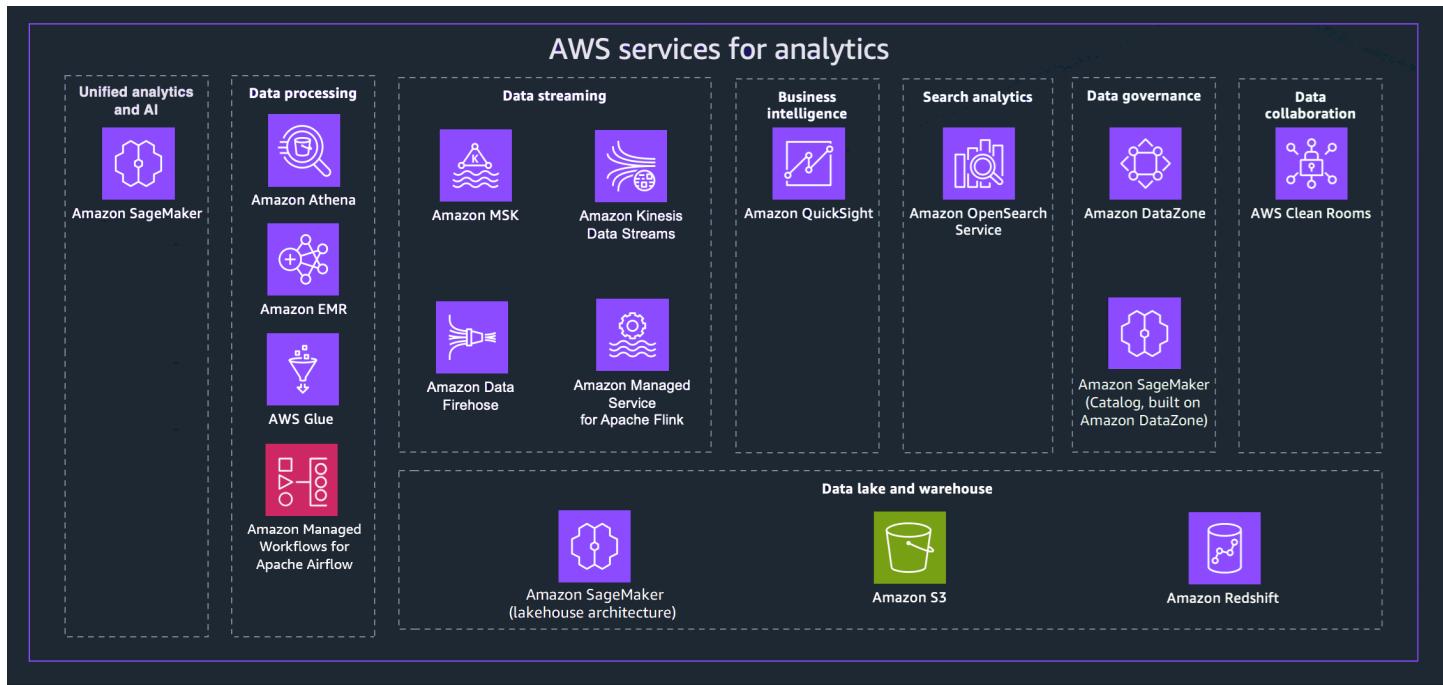
[이 8분 연설은 re:Invent 2024에서 Sirish Chandrasekaran과 Rick Sears가 1시간 동안 발표한 것입니다. 가상 회사인 Maxdome이 차세대 Amazon SageMaker의 일부인 SageMaker Unified Studio AI 및 분석을 사용하여 데이터 인사이트를 얻는 방법에 대한 개요를 제공합니다. Amazon SageMaker](#)

AWS 분석 서비스 이해

최신 데이터 전략은 데이터를 관리, 액세스, 분석 및 조치를 수행하는 데 도움이 되는 일련의 기술 구성 요소로 구축됩니다. 또한 데이터 소스에 연결할 수 있는 여러 옵션을 제공합니다. 최신 데이터 전략은 팀이 다음을 수행할 수 있도록 지원해야 합니다.

- 선호하는 도구 또는 기법 사용
- 인공 지능(AI)을 사용하여 데이터에 대한 특정 질문에 대한 답변을 찾는 데 도움이 됩니다.
- 적절한 보안 및 데이터 거버넌스 제어를 통해 데이터에 액세스할 수 있는 사용자 관리
- 데이터 파일로를 분석하여 데이터 레이크와 특별히 구축된 데이터 스토어를 모두 최대한 활용할 수 있습니다.
- 저렴한 비용으로 개방형 표준 기반 데이터 형식으로 원하는 양의 데이터 저장
- 데이터 레이크, 데이터 웨어하우스, 운영 데이터베이스, 애플리케이션 및 페더레이션 데이터 소스를 일관된 전체로 연결

AWS는 최신 데이터 전략을 달성하는 데 도움이 되는 다양한 서비스를 제공합니다. 다음 다이어그램은 이 가이드에서 다루는 분석 AWS 서비스를 보여줍니다. 다음 탭은 추가 세부 정보를 제공합니다.



Unified analytics and AI

차세대 [Amazon SageMaker](#)는 널리 채택된 AWS 기계 학습(ML)과 분석 기능을 결합하여 분석 및 AI에 대한 통합 경험을 제공하고 모든 데이터에 대한 통합 액세스를 제공합니다. [Amazon SageMaker Unified Studio](#)를 사용하면 소프트웨어 개발을 위한 생성형 AI 어시스턴트인 Amazon Q Developer에 의해 가속화된 모델 개발, 생성형 AI 애플리케이션 개발, 데이터 처리 및 SQL 분석을 위한 친숙한 AWS 도구를 사용하여 더 빠르게 협업하고 구축할 수 있습니다. 엔터프라이즈 보안 요구 사항을 충족하기 위해 내장된 거버넌스를 사용하여 데이터 레이크, 데이터 웨어하우스 또는 타사 및 페더레이션 소스에서 데이터에 액세스합니다.

Data processing

- [Amazon Athena](#)는 Amazon S3에 저장된 비정형, 반정형 및 정형 데이터를 분석하는 데 도움이 됩니다. 예를 들면 CSV, JSON 또는 컬럼 방식 데이터 형식(예: Apache Parquet 및 Apache ORC)이 해당됩니다. Athena를 사용하면 데이터를 집계하거나 Athena로 로드할 필요 없이 ANSI SQL을 사용한 임의 쿼리를 실행할 수 있습니다. Athena는 Quick Suite AWS Glue Data Catalog 및 [기타 AWS 서비스](#)와 통합됩니다. 인프라를 관리할 필요 없이 [Trino](#)를 사용하여 대규모로 데이터를 분석하고 Apache Flink 및 Apache Spark를 사용하여 실시간 분석을 구축할 수도 있습니다.
- [Amazon EMR](#)은 Apache Hadoop 및 Apache Spark와 같은 빅 데이터 프레임워크 실행을 간소화 AWS 하여 방대한 양의 데이터를 처리하고 분석하는 관리형 클러스터 플랫폼입니다. 이러한 프레임워크와 함께 관련 오픈 소스 프로젝트를 사용하여 분석용 데이터와 비즈니스 인텔리전스 워크로드를 처리할 수 있습니다. 또한 Amazon EMR을 사용하면 Amazon S3와 같은 다른 데이터 스토어 및 데이터베이스 안팎으로 대량의 AWS 데이터를 변환하고 이동할 수 있습니다.

- 를 사용하면 100개 이상의 다양한 데이터 소스를 검색 및 연결하고 중앙 집중식 데이터 카탈로그에서 데이터를 관리할 [AWS Glue](#) 수 있습니다. ETL 파이프라인을 시각적으로 생성, 실행 및 모니터링하여 데이터를 데이터 레이크로 로드할 수 있습니다. 또한 Athena, Amazon EMR 및 Amazon Redshift Spectrum을 사용하여 카탈로그화된 데이터를 즉시 검색하고 쿼리할 수 있습니다.
- [Amazon Managed Workflows for Apache Airflow](#)(MWAA)는 클라우드에서 데이터 워크플로를 더 쉽게 생성, 예약 및 모니터링할 수 있는 Apache Airflow의 완전 관리형 구현입니다. MWAA는 필요에 맞게 워크플로 용량을 자동으로 조정하고 보안 서비스와 통합합니다 AWS . MWAA를 사용하여 데이터 처리, ETL 작업, 기계 학습 파이프라인을 포함한 분석 서비스 전반에서 워크플로를 오키스트레이션할 수 있습니다.

Data streaming

- [Amazon Managed Streaming for Apache Kafka](#)(Amazon MSK)를 사용하면 Apache Kafka를 사용하여 스트리밍 데이터를 처리하는 애플리케이션을 구축하고 실행할 수 있습니다. Amazon MSK는 클러스터 생성, 업데이트, 삭제와 같은 컨트롤 플레인 작업을 제공합니다. 따라서 데이터 생성 및 소비와 같은 Apache Kafka 데이터 영역 작업을 사용할 수 있습니다.
- [Amazon Kinesis Data Streams](#)를 사용하면 대규모 데이터 레코드 스트림을 실시간으로 수집하고 처리할 수 있습니다. 사용되는 데이터 유형으로는 IT 인프라 로그 데이터, 애플리케이션 로그, 소셜 미디어, 시장 데이터 피드, 웹 클릭스트림 데이터 등이 있습니다.
- [Amazon Data Firehose](#)는 Amazon S3, Amazon Redshift, Amazon OpenSearch Service, Splunk, Apache Iceberg Tables 등의 대상으로 실시간 스트리밍 데이터를 제공하기 위한 완전관리형 서비스입니다. Datadog, Dynatrace, LogicMonitor, MongoDB, New Relic, Coralogix, Elastic 등 지원되는 타사 서비스 공급자가 소유한 모든 사용자 지정 HTTP 엔드포인트 또는 HTTPS 엔드포인트로 데이터를 전송할 수도 있습니다.
- [Amazon Managed Service for Apache Flink](#)를 사용하면 Java, Scala, Python 또는 SQL을 사용하여 스트리밍 데이터를 처리하고 분석할 수 있습니다. 스트리밍 소스 및 정적 소스에 대해 코드를 작성하고 실행하여 시계열 분석을 수행하고 실시간 대시보드 및 지표를 제공할 수 있습니다.

Business intelligence

[Quick Suite](#)는 의사 결정자에게 대화형 시각적 환경에서 정보를 탐색하고 해석할 수 있는 기회를 제공합니다. 단일 데이터 대시보드에서 Quick Suite에는 AWS 데이터, 타사 데이터, 빅 데이터, 스프레드시트 데이터, SaaS 데이터, B2B 데이터 등이 포함될 수 있습니다. Quick Suite Q를 사용하면 자연어를 사용하여 데이터에 대해 질문하고 응답을 받을 수 있습니다. 예: "캘리포니아에서 가장 많이 팔리는 범주는 무엇입니까?"

Search analytics

[Amazon OpenSearch Service](#)는 OpenSearch 클러스터에 대한 모든 리소스를 프로비저닝하고 시작합니다. 또한 실패한 OpenSearch Service 노드를 자동으로 감지한 다음 교체해 자체 관리형 인프라와 관련된 오버헤드를 줄입니다. OpenSearch Service 디렉트 쿼리를 사용하여 Amazon S3 및 기타 AWS 서비스의 데이터를 분석할 수 있습니다.

Data governance

[Amazon DataZone](#)을 사용하면 세분화된 제어를 사용하여 데이터에 대한 액세스를 관리하고 관리할 수 있습니다. 이러한 제어는 적절한 수준의 권한 및 컨텍스트로 액세스를 보장하는 데 도움이 됩니다. Amazon DataZone은 Amazon Redshift, Athena, Quick Suite, AWS Glue온프레미스 소스 및 타사 소스를 포함한 데이터 관리 서비스를 통합하여 아키텍처를 간소화합니다.

Data collaboration

[AWS Clean Rooms](#)는 원시 데이터에 대한 액세스를 제공하지 않고도 집합 데이터 세트를 분석할 수 있는 안전한 공동 작업 영역입니다. 협업할 파트너를 선택하고, 데이터 세트를 선택하고, 해당 파트너를 위한 개인 정보 보호 강화 제어를 구성하여 다른 회사와 협업할 수 있습니다. 쿼리를 실행하면 해당 데이터의 원래 위치에서 데이터를 AWS Clean Rooms 읽고 기본 제공 분석 규칙을 적용하여 해당 데이터에 대한 제어를 유지하는 데 도움이 됩니다.

Data lake and data warehouse

- [차세대 Amazon SageMaker](#)는 Apache Iceberg와 완벽하게 호환되므로 Amazon Simple Storage Service(Amazon S3) 데이터 레이크와 Amazon Redshift 데이터 웨어하우스에서 데이터를 통합할 수 있습니다. 이를 통해 단일 데이터 사본에서 분석 및 AI 및 기계 학습(ML) 애플리케이션을 구축할 수 있습니다. 제로 ETL 통합을 통해 운영 소스의 데이터를 거의 실시간으로 스트리밍하고, 여러 소스에서 페더레이션 쿼리를 실행하고, Apache Iceberg 호환 도구를 사용하여 데이터에 액세스할 수 있습니다. 모든 분석 및 ML 도구 및 엔진에 적용되는 세분화된 권한을 정의하여 데이터를 보호할 수 있습니다.
- [Amazon S3](#)는 데이터 레이크 기반에 사용할 수 있는 거의 모든 양의 데이터를 저장하고 보호할 수 있습니다. Amazon S3는 특정 비즈니스, 조직 및 규정 준수 요구 사항에 맞게 데이터에 대한 액세스를 최적화, 구조화 및 구성할 수 있는 관리 기능을 제공합니다. Amazon S3 Tables는 분석 워크로드에 최적화된 S3 스토리지를 제공합니다. 표준 SQL 문을 사용하여 Athena, Amazon Redshift, Apache Spark와 같은 Iceberg를 지원하는 쿼리 엔진으로 테이블을 쿼리할 수 있습니다.
- [Amazon Redshift](#)는 완전 관리형 페타바이트 규모의 데이터 웨어하우스 서비스입니다. Amazon Redshift는 Amazon SageMaker의 데이터 레이크하우스에 연결할 수 있으므로 Amazon Redshift 데이터 웨어하우스 및 Amazon S3 데이터 레이크 전반의 통합 데이터에 강력한 SQL 분석 기능을

사용할 수 있습니다. 자연어를 통해 SQL 작성률을 간소화하는 Amazon Redshift에서 Amazon Q를 사용할 수도 있습니다.

AWS 분석 서비스의 기준 고려

데이터 분석을 구축하는 데는 여러 가지 이유가 있습니다 AWS. 클라우드 마이그레이션 여정의 첫 단계로 그린필드 또는 파일럿 프로젝트를 지원해야 할 수 있습니다. 또는 중단을 최소화하면서 기존 워크로드를 마이그레이션할 수도 있습니다. 어떤 목표를 선택하든 다음과 같은 고려 사항이 유용할 수 있습니다.

Assess data sources and data types

사용 가능한 데이터 소스 및 데이터 유형을 분석하여 데이터 다양성, 빈도 및 품질을 포괄적으로 이해합니다. 데이터 처리 및 분석과 관련된 잠재적 문제를 이해합니다. 이 분석은 다음과 같은 이유로 매우 중요합니다.

- 데이터 소스는 다양하며 다양한 시스템, 애플리케이션, 디바이스 및 외부 플랫폼에서 가져옵니다.
- 데이터 소스에는 고유한 구조, 형식 및 데이터 업데이트 빈도가 있습니다. 이러한 소스를 분석하면 적절한 데이터 수집 방법 및 기술을 식별하는 데 도움이 됩니다.
- 정형, 반정형 및 비정형 데이터와 같은 데이터 유형을 분석하면 적절한 데이터 처리 및 스토리지 접근 방식이 결정됩니다.
- 데이터 소스 및 유형을 분석하면 데이터 품질 평가를 용이하게 하고 값 누락, 불일치 또는 부정확성 등 잠재적 데이터 품질 문제를 예측할 수 있습니다.

Data processing requirements

데이터를 수집, 변환, 정리 및 분석 준비하는 방법에 대한 데이터 처리 요구 사항을 결정합니다. 주요 고려 사항은 다음과 같습니다.

- 데이터 변환: 원시 데이터를 분석에 적합하게 만드는데 필요한 특정 변환을 결정합니다. 여기에는 데이터 집계, 정규화, 필터링 및 보강과 같은 작업이 포함됩니다.
- 데이터 정리: 데이터 품질을 평가하고 누락되거나 부정확하거나 일관되지 않은 데이터를 처리하는 프로세스를 정의합니다. 데이터 정리 기술을 구현하여 신뢰할 수 있는 인사이트를 위한 고품질 데이터를 보장합니다.

- 처리 빈도: 분석 요구 사항에 따라 실시간, 거의 실시간 또는 배치 처리가 필요한지 여부를 결정합니다. 실시간 처리는 즉각적인 인사이트를 제공하는 반면 배치 처리는 주기적 분석에 충분할 수 있습니다.
- 확장성 및 처리량: 데이터 볼륨, 처리 속도 및 동시 데이터 요청 수를 처리하기 위한 확장성 요구 사항을 평가합니다. 선택한 처리 접근 방식이 향후 성장을 수용할 수 있는지 확인합니다.
- 지연 시간: 데이터 처리에 허용되는 지연 시간과 데이터 수집부터 분석 결과까지 걸리는 시간을 고려합니다. 이는 실시간 또는 시간에 민감한 분석에 특히 중요합니다.

Storage requirements

분석 파이프라인 전체에서 데이터가 저장되는 방식과 위치를 결정하여 스토리지 요구 사항을 결정합니다. 중요한 고려 사항은 다음과 같습니다.

- 데이터 볼륨: 생성 및 수집되는 데이터의 양을 평가하고 향후 데이터 증가를 예측하여 충분한 스토리지 용량을 계획합니다.
- 데이터 보존: 기록 분석 또는 규정 준수를 위해 데이터를 보존해야 하는 기간을 정의합니다. 적절한 데이터 보존 정책을 결정합니다.
- 데이터 액세스 패턴: 가장 적합한 스토리지 솔루션을 선택하기 위해 데이터에 액세스하고 쿼리하는 방법을 이해합니다. 읽기 및 쓰기 작업, 데이터 액세스 빈도 및 데이터 로캘을 고려합니다.
- 데이터 보안: 민감한 정보를 보호하기 위해 암호화 옵션, 액세스 제어 및 데이터 보호 메커니즘을 평가하여 데이터 보안의 우선순위를 지정합니다.
- 비용 최적화: 데이터 액세스 패턴 및 사용량을 기반으로 가장 비용 효율적인 스토리지 솔루션을 선택하여 스토리지 비용을 최적화합니다.
- 분석 서비스와의 통합: 선택한 스토리지 솔루션과 파이프라인의 데이터 처리 및 분석 도구 간의 원활한 통합을 보장합니다.

Types of data

데이터 수집 및 수집을 위한 분석 서비스를 결정할 때는 조직의 요구 사항 및 목표와 관련된 다양한 유형의 데이터를 고려하세요. 고려해야 할 수 있는 일반적인 데이터 유형은 다음과 같습니다.

- 트랜잭션 데이터: 고객 구매, 금융 거래, 온라인 주문, 사용자 활동 로그와 같은 개별 상호 작용 또는 거래에 대한 정보를 포함합니다.
- 파일 기반 데이터: 로그 파일, 스프레드시트, 문서, 이미지, 오디오 파일 및 비디오 파일과 같이 파일에 저장된 정형 또는 비정형 데이터를 나타냅니다. 분석 서비스는 다양한 파일 형식의 수집을 지원해야 합니다.

- 이벤트 데이터: 사용자 작업, 시스템 이벤트, 시스템 이벤트 또는 비즈니스 이벤트와 같은 중요한 발생 또는 인시던트를 캡처합니다. 이벤트에는 온스트림 또는 다운스트림 처리를 위해 캡처되는 고속으로 도착하는 모든 데이터가 포함될 수 있습니다.

Operational considerations

운영 책임은 사용자와 간에 공유되며 AWS, 책임 분담은 현대화 수준에 따라 달라집니다.에서 분석 인프라를 자체 관리 AWS하거나 다양한 서비스 분석 서비스를 활용하여 인프라 관리 부담을 줄일 수 있습니다.

자체 관리형 옵션을 사용하면 인프라 및 구성의 더 잘 제어할 수 있지만 더 많은 운영 노력이 필요합니다.

서비스 옵션은 대부분의 운영 부담을 추상화하여 자동 확장성, 고가용성 및 강력한 보안 기능을 제공하므로 사용자는 인프라 및 운영 작업을 관리하는 대신 분석 솔루션을 구축하고 인사이트를 도출하는 데 더 집중할 수 있습니다. 서비스 분석 솔루션의 이점은 다음과 같습니다.

- 인프라 추상화: 서비스 서비스는 인프라 관리를 추상화하여 사용자가 프로비저닝, 규모 조정 및 유지 관리 작업을 수행하지 못하도록 합니다. 이러한 운영 측면을 AWS 처리하여 관리 오버헤드를 줄입니다.
- 자동 규모 조정 및 성능: 서비스 서비스는 워크로드 수요에 따라 리소스를 자동으로 규모 조정하여 수동 개입 없이 최적의 성능을 보장합니다.
- 고가용성 및 재해 복구: AWS는 서비스 서비스에 고가용성을 제공합니다. 데이터 중복성, 복제 및 재해 복구를 AWS 관리하여 데이터 가용성과 신뢰성을 향상시킵니다.
- 보안 및 규정 준수: AWS 업계 표준 및 모범 사례를 준수하면서 서비스 서비스에 대한 보안 조치, 데이터 암호화 및 규정 준수를 관리합니다.
- 모니터링 및 로깅: AWS 서비스 서비스에 내장된 모니터링, 로깅 및 알림 기능을 제공합니다. 사용자는 Amazon CloudWatch를 통해 세부 지표 및 로그에 액세스할 수 있습니다.

Type of workload

최신 분석 파이프라인을 구축할 때 다양한 분석 요구 사항을 효과적으로 충족하려면 지원할 워크로드 유형을 결정하는 것이 중요합니다. 각 워크로드 유형에 대해 고려해야 할 주요 결정 사항은 다음과 같습니다.

배치 워크로드

- 데이터 볼륨 및 빈도: 배치 처리는 정기 업데이트가 있는 대용량 데이터에 적합합니다.

- 데이터 지연 시간: 배치 처리로 인해 실시간 처리에 비해 인사이트 제공이 다소 지연될 수 있습니다.

대화형 분석

- 데이터 쿼리 복잡성: 대화형 분석에는 빠른 피드백을 위해 지연 시간이 짧은 응답이 필요합니다.
- 데이터 시각화: 비즈니스 사용자가 데이터를 시각적으로 탐색할 수 있도록 대화형 데이터 시각화 도구의 필요성을 평가합니다.

스트리밍 워크로드

- 데이터 속도 및 볼륨: 스트리밍 워크로드는 고속 데이터를 처리하기 위해 실시간 처리가 필요합니다.
- 데이터 윈도우 설정: 관련 인사이트를 추출하기 위해 스트리밍 데이터에 대한 데이터 윈도우 설정 및 시간 기반 집계를 정의합니다.

Type of analysis needed

분석으로부터 도출하려는 비즈니스 목표와 인사이트를 명확하게 정의합니다. 분석 유형에 따라 용도가 다릅니다. 예:

- 설명 분석은 기록 개요를 얻는 데 이상적입니다.
- 진단 분석은 과거 이벤트의 이유를 이해하는 데 도움이 됩니다.
- 예측 분석은 미래 결과를 예측합니다.
- 권장 분석은 최적의 작업에 대한 권장 사항을 제공합니다.

비즈니스 목표를 관련 분석 유형과 일치시킵니다. 다음은 올바른 유형의 분석을 선택하는 데 도움이 되는 몇 가지 주요 결정 기준입니다.

- 데이터 가용성 및 품질: 설명 및 진단 분석은 과거 데이터에 의존하는 반면 예측 및 규범적 분석에는 정확한 모델을 구축하기 위한 충분한 과거 데이터와 고품질 데이터가 필요합니다.
- 데이터 볼륨 및 복잡성: 예측 및 규범적 분석에는 상당한 데이터 처리와 컴퓨팅 리소스가 필요합니다. 인프라와 도구가 데이터 볼륨과 복잡성을 처리할 수 있는지 확인합니다.
- 결정 복잡성: 결정에 여러 변수, 제약 조건 및 목표가 포함된 경우 최적의 조치를 안내하는 데 규범적 분석이 더 적합할 수 있습니다.

- 위험 허용 범위: 규범적 분석은 권장 사항을 제공할 수 있지만 관련 불확실성이 수반됩니다. 의사 결정자가 분석 출력과 관련된 위험을 이해해야 합니다.

Evaluate scalability and performance

아키텍처의 확장성 및 성능 요구 사항을 평가합니다. 설계는 증가하는 데이터 볼륨, 사용자 요구 및 분석 워크로드를 처리해야 합니다. 고려해야 할 주요 결정 요인은 다음과 같습니다.

- 데이터 볼륨 및 성장: 현재 데이터 볼륨을 평가하고 향후 성장을 예상합니다.
- 데이터 속도 및 실시간 요구 사항: 데이터를 실시간으로 또는 거의 실시간으로 처리하고 분석해야 하는지 결정합니다.
- 데이터 처리 복잡성: 데이터 처리 및 분석 작업의 복잡성을 분석합니다. 컴퓨팅 집약적인 작업의 경우 Amazon EMR과 같은 서비스는 빅 데이터 처리를 위한 확장 가능하고 관리형 환경을 제공합니다.
- 동시성 및 사용자 로드: 동시 사용자 수와 시스템의 사용자 로드 수준을 고려합니다.
- Auto Scaling 기능: 수요에 따라 리소스가 자동으로 확장 또는 축소될 수 있도록 Auto Scaling 기능을 제공하는 서비스를 고려합니다. 이렇게 하면 효율적인 리소스 사용률과 비용 최적화가 보장됩니다.
- 지리적 배포: 데이터 아키텍처를 여러 리전 또는 위치에 배포해야 하는 경우 글로벌 복제 및 지연 시간이 짧은 데이터 액세스가 가능한 서비스를 고려합니다.
- 비용 대비 성능 절충: 성능 요구 사항과 비용 고려 사항의 균형을 맞춥니다. 성능이 뛰어난 서비스는 비용이 더 많이 들 수 있습니다.
- 서비스 수준 계약(SLAs): AWS 서비스에서 제공하는 SLAs를 확인하여 확장성 및 성능 기대치를 충족하는지 확인합니다.

Data governance

데이터 거버넌스는 데이터 자산의 효과적인 관리, 품질, 보안 및 규정 준수를 보장하기 위해 구현해야 하는 프로세스, 정책 및 제어 세트입니다. 고려해야 할 주요 결정 사항은 다음과 같습니다.

- 데이터 보존 정책: 규제 요구 사항 및 비즈니스 요구 사항에 따라 데이터 보존 정책을 정의하고 더 이상 필요하지 않은 경우 안전한 데이터 폐기를 위한 프로세스를 설정합니다.
- 감사 추적 및 로깅: 로깅 및 감사 메커니즘을 결정하여 데이터 액세스 및 사용량을 모니터링합니다. 포괄적인 감사 추적을 구현하여 규정 준수 및 보안 모니터링을 위한 데이터 변경, 액세스 시도 및 사용자 활동을 추적합니다.

- 규정 준수 요구 사항: 조직에 적용되는 산업별 및 지리적 데이터 규정 준수 규정을 이해합니다. 데이터 아키텍처가 이러한 규정 및 지침에 부합하는지 확인합니다.
- 데이터 분류: 민감도에 따라 데이터를 분류하고 각 데이터 클래스에 적합한 보안 제어를 정의합니다.
- 재해 복구 및 비즈니스 연속성: 예기치 않은 이벤트 또는 시스템 장애 발생 시 데이터 가용성과 복원력을 보장하기 위해 재해 복구 및 비즈니스 연속성을 계획합니다.
- 타사 데이터 공유: 타사 엔터티와 데이터를 공유하는 경우 보안 데이터 공유 프로토콜 및 계약을 구현하여 데이터 기밀성을 보호하고 데이터 오용을 방지합니다.

Security

분석 파이프라인의 데이터 보안에는 파이프라인의 모든 단계에서 데이터를 보호하여 기밀성, 무결성 및 가용성을 보장하는 것이 포함됩니다. 고려해야 할 주요 결정 사항은 다음과 같습니다.

- 액세스 제어 및 권한 부여: 강력한 인증 및 권한 부여 프로토콜을 구현하여 권한이 있는 사용자만 특정 데이터 리소스에 액세스할 수 있도록 합니다.
- 데이터 암호화: 데이터베이스, 데이터 레이크에 저장된 데이터와 아키텍처의 다양한 구성 요소 간 데이터 이동 중에 적절한 암호화 방법을 선택합니다.
- 데이터 마스킹 및 익명화: 특정 분석 프로세스를 계속하면서 PII 또는 민감한 비즈니스 데이터와 같은 민감한 데이터를 보호하기 위해 데이터 마스킹 또는 익명화의 필요성을 고려합니다.
- 보안 데이터 통합: 데이터 이동 중에 데이터 유출 또는 무단 액세스를 방지하여 아키텍처의 다양한 구성 요소 간에 데이터가 안전하게 흐르도록 보안 데이터 통합 관행을 수립합니다.
- 네트워크 격리: 리소스를 퍼블릭 인터넷에 노출하지 않도록 [Amazon VPC 엔드포인트](#)를 지원하는 서비스를 고려합니다.

Plan for integration and data flows

분석 파이프라인의 다양한 구성 요소 간의 통합 지점과 데이터 흐름을 정의하여 원활한 데이터 흐름과 상호 운용성을 보장합니다. 고려해야 할 주요 결정 사항은 다음과 같습니다.

- 데이터 소스 통합: 데이터베이스, 애플리케이션, 파일 또는 외부 APIs. 데이터 수집 방법(배치, 실시간, 이벤트 기반)을 결정하여 지연 시간을 최소화하면서 효율적으로 파이프라인으로 데이터를 가져옵니다.
- 데이터 변환: 분석을 위한 데이터를 준비하는 데 필요한 변환을 결정합니다. 파이프라인을 통해 이동하는 데이터를 정리, 집계, 정규화 또는 보강할 도구와 프로세스를 결정합니다.

- 데이터 이동 아키텍처: 파이프라인 구성 요소 간의 데이터 이동에 적합한 아키텍처를 선택합니다. 실시간 요구 사항과 데이터 볼륨에 따라 배치 처리, 스트림 처리 또는 둘의 조합을 고려합니다.
- 데이터 복제 및 동기화: 모든 구성 요소에서 데이터를 up-to-date 유지하기 위해 데이터 복제 및 동기화 메커니즘을 결정합니다. 데이터 최신성 요구 사항에 따라 실시간 복제 솔루션 또는 주기적 데이터 동기화를 고려합니다.
- 데이터 품질 및 검증: 데이터 품질 확인 및 검증 단계를 구현하여 파이프라인을 통해 이동하는 데이터의 무결성을 보장합니다. 알림 또는 오류 처리와 같이 데이터 검증에 실패할 때 수행할 작업을 결정합니다.
- 데이터 보안 및 암호화: 전송 중 및 저장 시 데이터를 보호하는 방법을 결정합니다. 데이터 민감도에 따라 필요한 보안 수준을 고려하여 파이프라인 전체에서 민감한 데이터를 보호하기 위한 암호화 방법을 결정합니다.
- 확장성 및 복원력: 데이터 흐름 설계가 수평 확장성을 허용하고 늘어난 데이터 볼륨과 트래픽을 처리할 수 있는지 확인합니다.

Architect for cost optimization

에서 분석 파이프라인을 구축하면 다양한 비용 최적화 기회가 AWS 제공됩니다. 비용 효율성을 보장하려면 다음 전략을 고려하세요.

- 리소스 크기 조정 및 선택: 실제 워크로드 요구 사항에 따라 리소스의 크기를 조정합니다. 오버프로비저닝을 방지하면서 워크로드 성능 요구 사항에 맞는 AWS 서비스 및 인스턴스 유형을 선택합니다.
- Auto Scaling: 다양한 워크로드를 경험하는 서비스에 Auto Scaling을 구현합니다. Auto Scaling은 수요에 따라 인스턴스 수를 동적으로 조정하여 트래픽이 적은 기간 동안 비용을 절감합니다.
- 스팟 인스턴스: 중요하지 않고 내결함성이 뛰어난 워크로드에 Amazon EC2 스팟 인스턴스를 사용합니다. 스팟 인스턴스는 온디맨드 인스턴스에 비해 비용을 크게 절감할 수 있습니다.
- 예약 인스턴스: 예측 가능한 사용량으로 안정적인 워크로드에 대한 온디맨드 요금보다 상당한 비용 절감을 달성하려면 예약 인스턴스 구매를 AWS 고려하세요.
- 데이터 스토리지 계층화: 데이터 액세스 빈도에 따라 다양한 스토리지 클래스를 사용하여 데이터 스토리지 비용을 최적화합니다.
- 데이터 수명 주기 정책: 수명 및 사용 패턴에 따라 데이터를 자동으로 이동하거나 삭제하도록 데이터 수명 주기 정책을 설정합니다. 이를 통해 스토리지 비용을 관리하고 데이터 스토리지를 가치에 맞게 조정할 수 있습니다.

AWS 분석 서비스 선택

이제 분석 요구 사항을 평가하는 기준을 알았으므로 조직 요구 사항에 적합한 AWS 분석 서비스를 선택할 준비가 되었습니다. 다음 표는 공통 기능 및 비즈니스 목표에 맞게 서비스 세트를 정렬합니다.

Categories	무엇에 최적화되어 있나요?	서비스
통합 분석 및 AI	<p>분석 및 AI 개발</p> <p>단일 개발 환경인 Amazon SageMaker Unified Studio를 사용하여 데이터, 분석 및 AI 기능에 액세스하도록 최적화되었습니다.</p>	Amazon SageMaker
데이터 처리	<p>대화형 분석</p> <p>실시간 데이터 분석 및 탐색을 수행하도록 최적화되어 사용자가 대화형으로 데이터를 쿼리하고 시각화할 수 있습니다.</p>	Amazon Athena
	<p>빅 데이터 처리</p> <p>대량의 데이터를 처리, 이동 및 변환하는 데 최적화되었습니다.</p>	Amazon EMR
	<p>데이터 카탈로그</p> <p>사용 가능한 데이터, 구조, 특성 및 관계에 대한 자세한 정보를 제공하는 데 최적화되었습니다.</p>	AWS Glue
	<p>워크플로 오케스트레이션</p> <p>Apache Airflow를 사용하여 분석 프로세스 및 ETL 작업을 조</p>	Amazon MWAA

Categories	무엇에 최적화되어 있나요?	서비스
	정하는 데이터 워크플로를 생성, 예약 및 모니터링하는 데 최적화되었습니다.	
데이터 스트리밍	<p>스트리밍 데이터의 Apache Kafka 처리</p> <p>Apache Kafka 데이터 영역 작업을 사용하고 Apache Kafka의 오픈 소스 버전을 실행하는 데 최적화되었습니다.</p>	Amazon MSK
	실시간 처리	Amazon Kinesis Data Streams
	IT 인프라 로그 데이터, 애플리케이션 로그, 소셜 미디어, 시장 데이터 피드, 웹 클릭스트림 데이터를 포함한 빠르고 지속적인 데이터 수집 및 집계에 최적화되었습니다.	
	실시간 스트리밍 데이터 전송	Amazon Data Firehose
	Amazon S3, Amazon Redshift, OpenSearch Service, Splunk, Apache Iceberg Tables 및 지원되는 타사 서비스 공급자가 소유한 사용자 지정 HTTP 엔드포인트 또는 HTTP 엔드포인트와 같은 대상으로 실시간 스트리밍 데이터를 전송하는 데 최적화되었습니다.	

Categories	무엇에 최적화되어 있나요?	서비스
	<p>Apache Flink 애플리케이션 빌드</p> <p>Java, Scala, Python 또는 SQL을 사용하여 스트리밍 데이터를 처리하고 분석하도록 최적화되었습니다.</p>	Amazon Managed Service for Apache Flink
비즈니스 인텔리전스	<p>대시보드 및 시각화</p> <p>복잡한 데이터 세트를 시각적으로 표현하고 데이터에 대한 자연어 쿼리를 제공하는데 최적화되었습니다.</p>	Quick Suite
검색 분석	<p>관리형 OpenSearch 클러스터</p> <p>로그 분석, 실시간 애플리케이션 모니터링 및 클릭스트림 분석에 최적화되었습니다.</p>	Amazon OpenSearch Service
데이터 거버넌스	<p>데이터 액세스 관리</p> <p>수명 주기 동안 데이터의 적절한 관리, 가용성, 사용성, 무결성 및 보안을 설정하는데 최적화되었습니다.</p>	Amazon DataZone
데이터 공동 작업	<p>보안 데이터 클린룸</p> <p>원시 기본 데이터를 공유하지 않고 다른 회사와 협업하도록 최적화되었습니다.</p>	AWS Clean Rooms

Categories	무엇에 최적화되어 있나요?	서비스
데이터 레이크 및 웨어하우스	<p>데이터 레이크 및 데이터 웨어하우스에 대한 통합 액세스</p> <p>Amazon S3 데이터 레이크, Amazon Redshift 데이터 웨어하우스, 운영 데이터베이스, 타사 및 페더레이션 데이터 소스 전반에서 데이터 액세스를 통합하도록 최적화하는 레이크하우스 아키텍처를 기반으로 구축되었습니다.</p>	Amazon SageMaker
	<p>데이터 레이크용 객체 스토리지</p> <p>사실상 무제한의 확장성과 높은 내구성으로 데이터 레이크 기반을 제공하는 데 최적화되었습니다.</p>	Amazon S3
	<p>데이터 웨어하우징</p> <p>조직 내 다양한 소스에서 대량의 정형 및 반정형 데이터를 중앙에서 저장, 구성 및 검색하는데 최적화되었습니다.</p>	Amazon Redshift

AWS 분석 서비스 사용

이제 비즈니스 목표와 데이터 파이프라인 구축을 시작하기 위해 수집 및 분석할 데이터의 양과 속도를 명확하게 이해할 수 있습니다.

사용 방법을 탐색하고 사용 가능한 각 서비스에 대해 자세히 알아보기 위해 각 서비스의 작동 방식을 탐색하는 경로를 제공했습니다. 다음 섹션에서는 기본 사용부터 고급 심층 분석까지 시작할 수 있는 심층 설명서, 실습 자습서 및 리소스에 대한 링크를 제공합니다.

Amazon Athena

- Amazon Athena 시작하기

Amazon Athena를 사용하여 데이터를 쿼리하고 Amazon S3에 저장된 샘플 데이터를 기반으로 테이블을 생성하고, 테이블을 쿼리하고, 쿼리 결과를 확인하는 방법을 알아봅니다.

자습서 시작하기

- Athena에서 Apache Spark 시작하기

Athena 콘솔에서 간소화된 노트북 환경을 사용하여 Python 또는 Athena 노트북 APIs.

자습서 시작하기

- Amazon SageMaker 레이크하우스 아키텍처를 사용하여 Athena 페더레이션 쿼리 카탈로그 작성 및 관리

Amazon SageMaker의 데이터 레이크하우스를 통해 Amazon Redshift, DynamoDB 및 Snowflake에 저장된 데이터에 대한 페더레이션 쿼리를 연결, 관리 및 실행하는 방법을 알아봅니다.

블로그 읽기

- Athena를 사용하여 Amazon S3의 데이터 분석

사전 정의된 형식의 텍스트 파일로 생성된 Elastic Load Balancer의 로그에서 Athena를 사용하는 방법을 알아봅니다. 테이블을 생성하고, Athena에서 사용하는 형식으로 데이터를 분할하고, Parquet으로 변환하고, 쿼리 성능을 비교하는 방법을 보여줍니다.

블로그 게시물 읽기

AWS Clean Rooms

- 설정 AWS Clean Rooms

계정 AWS Clean Rooms에서 AWS를 설정하는 방법을 알아봅니다.

가이드 읽기

- 기본 데이터를 공유 AWS Clean Rooms 하지 않고에서 Entity Resolution을 사용하여 AWS 다자간 데이터 세트에서 데이터 인사이트 잠금 해제

준비 및 매칭을 사용하여 공동 작업자와의 데이터 매칭을 개선하는 방법을 알아봅니다.

블로그 게시물 읽기

- 차등 프라이버시가 개별 수준에서 데이터를 공개하지 않고 인사이트를 얻는 데 도움이 되는 방법

AWS Clean Rooms 차등 프라이버시가 차등 프라이버시 적용을 간소화하고 사용자의 프라이버시를 보호하는 데 어떻게 도움이 되는지 알아봅니다.

블로그 읽기

Amazon Data Firehose

- 자습서: 콘솔에서 Firehose 스트림 생성

AWS Management 콘솔 또는 AWS SDK를 사용하여 선택한 대상으로 Firehose 스트림을 생성하는 방법을 알아봅니다.

가이드 읽기

- Firehose 스트림으로 데이터 전송

다양한 데이터 소스를 사용하여 Firehose 스트림으로 데이터를 전송하는 방법을 알아봅니다.

가이드 읽기

- Firehose에서 소스 데이터 변환

Lambda 함수를 호출하여 수신 소스 데이터를 변환하고 변환된 데이터를 대상으로 전송하는 방법을 알아봅니다.

가이드 읽기

Amazon DataZone

- Amazon DataZone 시작하기

Amazon DataZone 루트 도메인을 생성하고, 데이터 포털 URL을 가져오고, 데이터 생산자 및 데이터 소비자를 위한 기본 Amazon DataZone 워크플로를 안내하는 방법을 알아봅니다.

자습서 시작하기

- 차세대 Amazon SageMaker 및 Amazon DataZone에서 데이터 계보의 일반 가용성 발표

Amazon DataZone이 자동 계보 캡처를 사용하여 AWS Glue 및 Amazon Redshift에서 계보 정보를 자동으로 수집하고 매핑하는 데 초점을 맞추는 방법을 알아봅니다.

블로그 읽기

Amazon EMR

- Amazon EMR 시작하기

Spark를 사용하여 샘플 클러스터를 시작하는 방법과 Amazon S3 버킷에 저장된 간단한 PySpark 스크립트를 실행하는 방법을 알아봅니다.

자습서 시작하기

- Amazon EKS에서 Amazon EMR 시작하기

가상 클러스터에 Spark 애플리케이션을 배포하여 Amazon EMR on Amazon EKS 사용을 시작하는 방법을 보여줍니다.

가이드 살펴보기

- EMR Serverless 시작하기

Amazon EMR Serverless가 최신 오픈 소스 프레임워크를 사용하는 분석 애플리케이션의 운영을 간소화하는 서비스 런타임 환경을 제공하는 방법을 알아봅니다.

자습서 시작하기

AWS Glue

- 시작하기 AWS Glue DataBrew

첫 번째 DataBrew 프로젝트를 생성하는 방법을 알아봅니다. 샘플 데이터 세트를 로드하고, 해당 데이터 세트에서 변환을 실행하고, 해당 변환을 캡처하는 레시피를 빌드하고, 변환된 데이터를 Amazon S3에 쓰는 작업을 실행합니다.

자습서 시작하기

- 를 사용하여 데이터 변환 AWS Glue DataBrew

데이터 분석가와 데이터 과학자가 데이터를 쉽게 정리하고 정규화하여 분석 및 기계 학습을 준비 할 수 있는 AWS Glue DataBrew 시각적 데이터 준비 도구에 대해 알아봅니다. 를 사용하여 ETL 프로세스를 구성하는 방법을 알아봅니다 AWS Glue DataBrew.

랩 시작하기

- AWS Glue DataBrew 침묵의 날

를 사용하여 분석 및 기계 학습 AWS Glue DataBrew 을 위한 데이터를 정리하고 정규화하는 방법을 알아봅니다.

워크숍 시작하기

- 시작하기 AWS Glue Data Catalog

Amazon S3 버킷을 데이터 소스로 AWS Glue Data Catalog 사용하는 첫 번째를 생성하는 방법을 알아봅니다.

자습서 시작하기

- 의 데이터 카탈로그 및 크롤러 AWS Glue

데이터 카탈로그의 정보를 사용하여 ETL 작업을 생성하고 모니터링하는 방법을 알아봅니다.

가이드 살펴보기

Amazon Kinesis Data Streams

- Amazon Kinesis Data Streams 자습서 시작하기

실시간 주식 데이터를 처리하고 분석하는 방법을 알아봅니다.

자습서 시작하기

- Amazon Kinesis Data Streams를 사용한 실시간 분석을 위한 아키텍처 패턴, 1부

시계열 데이터 분석과 이벤트 기반 마이크로서비스라는 두 가지 사용 사례의 일반적인 아키텍처 패턴에 대해 알아봅니다.

블로그 읽기

- Amazon Kinesis Data Streams를 사용한 실시간 분석을 위한 아키텍처 패턴, 2부

실시간 생성형 비즈니스 인텔리전스, 실시간 추천 시스템, 사물 인터넷 데이터 스트리밍 및 추론의 세 가지 시나리오에서 Kinesis Data Streams를 사용하는 AI 애플리케이션에 대해 알아봅니다.

블로그 읽기

Amazon Managed Service for Apache Flink

- Amazon Managed Service for Apache Flink란 무엇인가요?

Amazon Managed Service for Apache Flink의 기본 개념을 이해합니다.

가이드 살펴보기

- Amazon Managed Service for Apache Flink 워크숍

이 워크숍에서는 Amazon Managed Service for Apache Flink를 사용하여 Flink 애플리케이션을 배포, 운영 및 확장하는 방법을 알아봅니다.

가상 워크숍 참석

Amazon MSK

- Amazon MSK 시작하기

Amazon MSK 클러스터를 생성하고, 데이터를 생성 및 소비하고, 지표를 사용하여 클러스터의 상태를 모니터링하는 방법을 알아봅니다.

가이드 시작하기

- Amazon MSK 워크숍

이 실습용 Amazon MSK 워크숍을 자세히 살펴보세요.

워크숍 시작하기

Amazon MWAA

- Amazon MWAA 시작하기

첫 번째 MWAA 환경을 생성하고, Amazon S3에 DAG를 업로드하고, 첫 번째 워크플로를 실행하는 방법을 알아봅니다.

자습서 시작하기

- Amazon MWAA를 사용하여 데이터 파이프라인 구축

Glue, EMR 및 Redshift와 같은 다른 AWS 분석 서비스를 오케스트레이션하는 end-to-end 데이터 파이프라인을 구축하는 방법을 알아봅니다. 이 블로그 게시물에서는 Amazon Redshift에서 변환을 실행하는 작업과 함께 MWAA 및 Cosmos를 사용하여 dbt Core 작업을 오케스트레이션하는 간소화된 구성 기반 접근 방식을 살펴봅니다.

블로그 게시물 읽기

- Amazon MWAA 워크숍

실습을 살펴보고 데이터 워크플로 오케스트레이션에 Amazon MWAA를 배포, 구성 및 사용하는 방법을 알아봅니다.

워크숍 시작하기

- Amazon MWAA 모범 사례

분석 워크플로에서 Amazon MWAA를 사용하기 위한 아키텍처 패턴과 모범 사례를 알아봅니다.

가이드 읽기

OpenSearch Service

- OpenSearch Service 시작하기

Amazon OpenSearch Service를 사용하여 테스트 도메인을 생성하고 구성하는 방법을 알아봅니다.

자습서 시작하기

- OpenSearch Service 및 OpenSearch Dashboards를 사용하여 고객 지원 통화 시각화

기업이 몇 가지 고객 지원 전화를 받고 이를 분석하려는 상황을 전체적으로 살펴봅니다. 각 통화의 주제는 무엇입니까? 긍정적인 내용은 몇 통이었습니까? 부정적인 내용은 몇 통이었습니까? 관리자가 이러한 통화의 녹취록을 검색하거나 검토하려면 어떻게 해야 합니까?

자습서 시작하기

- Amazon OpenSearch Serverless 워크숍 시작하기

콘솔에서 AWS 새 Amazon OpenSearch Serverless 도메인을 설정하는 방법을 알아봅니다. 사용 가능한 다양한 유형의 검색 쿼리를 살펴보고, 매력적인 시각화를 설계하고, 할당된 사용자 권한을 기반으로 도메인과 문서를 보호하는 방법을 알아봅니다.

워크숍 시작하기

- **비용 최적화 벡터 데이터베이스: Amazon OpenSearch Service 양자화 기법 소개**

OpenSearch Service가 스칼라 및 제품 양자화 기술을 지원하여 메모리 사용량을 최적화하고 운영 비용을 절감하는 방법을 알아봅니다.

블로그 게시물 읽기

Quick Suite

- **Quick Suite 데이터 분석 시작하기**

첫 번째 분석을 생성하는 방법을 알아봅니다. 샘플 데이터를 사용하여 단순 분석 또는 고급 분석을 생성합니다. 또는 자체 데이터에 연결하여 분석을 생성할 수 있습니다.

가이드 살펴보기

- **Quick Suite를 사용한 시각화**

를 사용하여 비즈니스 인텔리전스(BI) 및 데이터 시각화의 기술적 측면을 알아봅니다 AWS. 대시보드를 애플리케이션 및 웹 사이트에 내장하고 액세스 및 권한을 안전하게 관리하는 방법을 알아봅니다.

과정 시작하기

- **Quick Suite 워크숍**

워크숍을 통해 Quick Suite 여정을 시작하세요.

워크숍 시작하기

Amazon Redshift

- **Amazon Redshift Serverless 시작하기**

Amazon Redshift Serverless의 기본 흐름을 이해하여 서비스 리소스를 생성하고, Amazon Redshift Serverless에 연결하고, 샘플 데이터를 로드한 다음 데이터에 대한 쿼리를 실행합니다.

가이드 살펴보기

- Amazon Redshift 심층 분석 워크숍

사용자가 Amazon Redshift 플랫폼 사용을 시작하는 데 도움이 되는 일련의 연습을 살펴봅니다.

워크숍 시작하기

Amazon S3

- Amazon S3 시작하기

첫 번째 DataBrew 프로젝트를 생성하는 방법을 알아봅니다. 샘플 데이터 세트를 로드하고, 해당 데이터 세트에서 변환을 실행하고, 해당 변환을 캡처하는 레시피를 빌드하고, 변환된 데이터를 Amazon S3에 쓰는 작업을 실행합니다.

가이드 시작하기

Amazon SageMaker

- SageMaker 시작하기

프로젝트를 생성하고, 구성원을 추가하고, 샘플 JupyterLab 노트북을 사용하여 빌드를 시작하는 방법을 알아봅니다.

가이드 읽기

- 차세대 Amazon SageMaker 소개: 모든 데이터, 분석 및 AI의 중심

데이터 처리, 모델 개발 및 생성형 AI 앱 개발을 시작하는 방법을 알아봅니다.

블로그 읽기

- SageMaker Unified Studio란 무엇인가요?

SageMaker Unified Studio의 기능과 Amazon SageMaker를 사용할 때 액세스하는 방법에 대해 알아봅니다.

가이드 읽기

- Amazon SageMaker의 레이크하우스 아키텍처 시작하기

Amazon SageMaker에서 프로젝트를 생성하고 비즈니스 사용 사례에 대한 데이터를 검색, 업로드 및 쿼리하는 방법을 알아봅니다.

가이드 읽기

- Amazon SageMaker의 레이크하우스 아키텍처의 데이터 연결

레이크하우스 아키텍처가 AWS 서비스 및 엔터프라이즈 애플리케이션 전반의 데이터 연결을 관리하는 통합 접근 방식을 제공하는 방법을 알아봅니다.

가이드 읽기

- SageMaker 레이크하우스 아키텍처를 사용하여 Athena 페더레이션 쿼리 카탈로그 작성 및 관리

Amazon SageMaker 프로젝트를 위해 Amazon Redshift, DynamoDB 및 Snowflake에 저장된 데이터에 대한 페더레이션 쿼리를 연결, 관리 및 실행하는 방법을 알아봅니다.

블로그 읽기

AWS 분석 서비스를 사용하는 방법 살펴보기

Editable architecture diagrams

참조 아키텍처 다이어그램

분석 솔루션을 개발, 확장 및 테스트하는 데 도움이 되는 아키텍처 다이어그램을 살펴보세요 AWS.

분석 참조 아키텍처 살펴보기

Ready-to-use code

추천 솔루션	AWS 솔루션
에서 Apache Druid를 사용한 확장 가능한 분석 AWS 비용 효율적이고 가용성이 높으며 복원력이 뛰어나고 내결합성이 뛰어난 호스팅 환경 AWS인 Apache Druid를 설정, 운영 및 관리하는 데 도움이	사전 구성되고 배포 가능한 솔루션과 이를 기반으로 구축된 구현 가이드를 살펴봅니다 AWS. 모든 AWS 보안, 자격 증명 및 거버넌스 솔루션 살펴보기

되는 배포 AWS빌드 코드입니다.

[이 솔루션 살펴보기](#)

Documentation

분석 백서

조직에 가장 적합한 분석 서비스를 선택, 구현 및 사용하는 방법에 대한 추가 인사이트와 모범 사례를 알아보려면 백서를 살펴보세요.

[분석 백서 살펴보기](#)

AWS 빅 데이터 블로그

특정 빅 데이터 사용 사례를 다루는 블로그 게시물을 살펴보세요.

[AWS 빅 데이터 블로그 살펴보기](#)

문서 기록

다음 표에서는 이 결정 가이드의 중요한 변경 사항을 설명합니다. 이 가이드 업데이트에 대한 알림을 받으려면 RSS 피드를 구독하면 됩니다.

변경 사항	설명	날짜
<u>re:Invent 업데이트</u>	결정 가이드 전반의 링크를 업데이트하고 Amazon Managed Workflows for Apache Airflow 를 추가했습니다.	2025년 9월 24일
<u>re:Invent 업데이트</u>	결정 가이드 전체에서 Amazon SageMaker, Amazon SageMaker Unified Studio(더 이상 미리 보기 없음) 및 Amazon SageMaker Lakehouse에 대한 참조가 업데이트되었습니다.	2025년 9월 9일
<u>re:Invent 업데이트</u>	SageMaker AI Unified Studio 및가 추가되었습니다 AWS Clean Rooms. 새로운 AI 기능으로 문서 전체가 업데이트되었습니다.	2025년 2월 20일
<u>최초 게시</u>	가이드가 처음 게시되었습니다.	2023년 11월 17일

기계 번역으로 제공되는 번역입니다. 제공된 번역과 원본 영어의 내용이 상충하는 경우에는 영어 버전이 우선합니다.