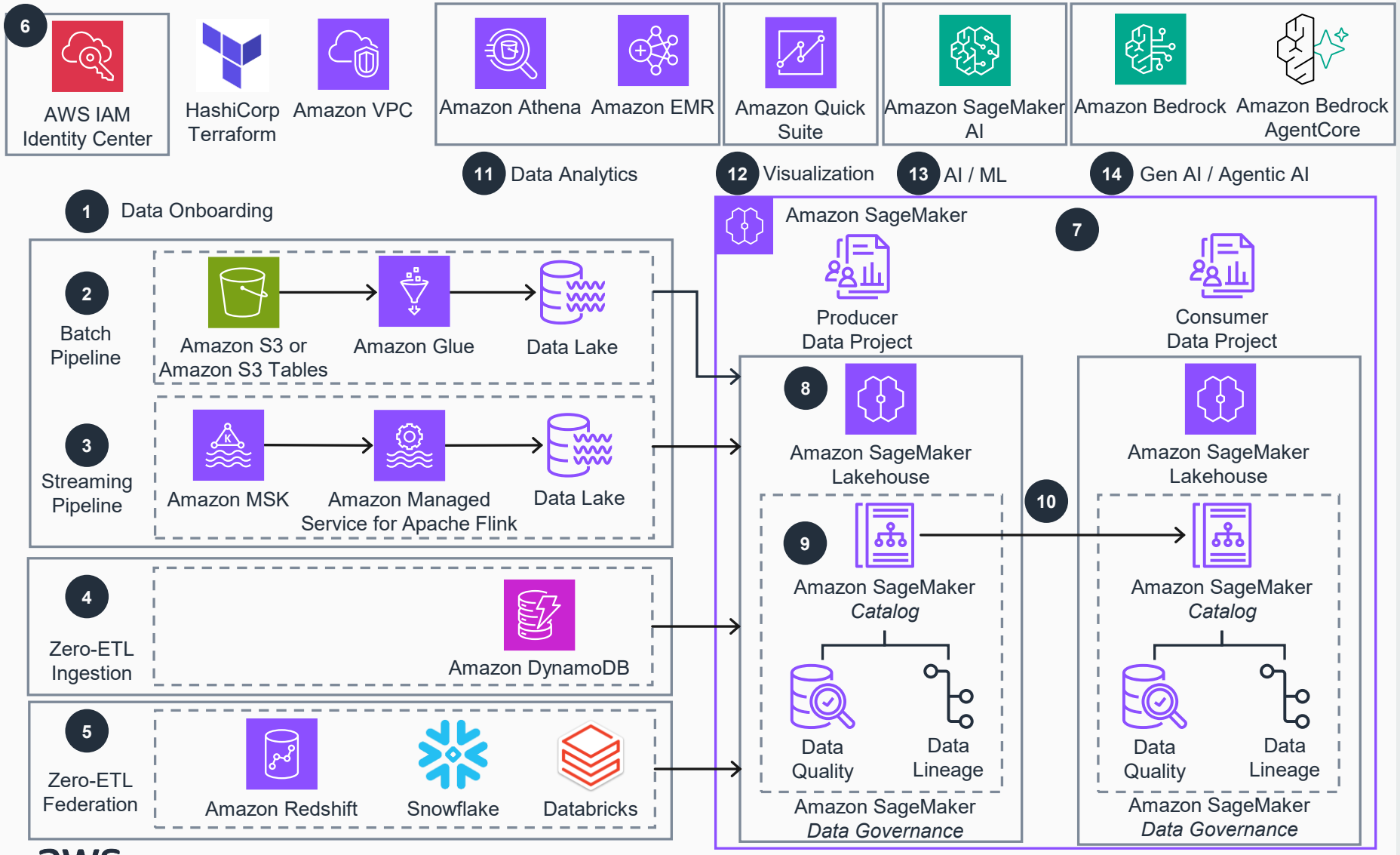
Guidance for Developing a Data & AI Foundation with Amazon SageMaker

This reference architecture illustrates how to rapidly deploy data, analytics, AI, and visualization solutions on AWS, using Amazon SageMaker, Amazon Quick Suite, and Amazon Bedrock AgentCore. This solution provides the Infrastructure-as-Code (IaC) building blocks to simplify the process of building enterprise data platforms on AWS. This slide shows steps 1-5.



AWS Reference Architecture

1 **Data Onboarding:** This solution provides IaC modules to provision various data onboarding mechanisms. Customers can implement batch pipelines from **Amazon Simple Storage Service (Amazon S3)** based data lakes, streaming pipelines, zero-ETL ingestion, and federated catalogs for external data sources.

2 **Batch Pipelines:** Batch data pipeline can be implemented using **AWS Glue** to ingest structured data from an **Amazon S3** bucket into hive or iceberg data lakes on **Amazon S3** bucket or **Amazon S3** table. Support includes: Hive data lakes on standard **Amazon S3** buckets 2) Iceberg data lakes on standard **Amazon S3** buckets and 3) Iceberg data lakes on **Amazon S3** table buckets. Alternately, use **Amazon EMR** for batch data pipelines.

3 **Streaming Pipelines:** Streaming data pipelines can be implemented using **Amazon Managed Service for Apache Flink** or **AWS Glue** Streaming to ingest messages from **Amazon Managed Streaming for Apache Kafka (Amazon MSK)** topics into hive or iceberg data lakes on an **Amazon S3** bucket or S3 table. The Flink option is particularly useful for performing Streaming Analytics.

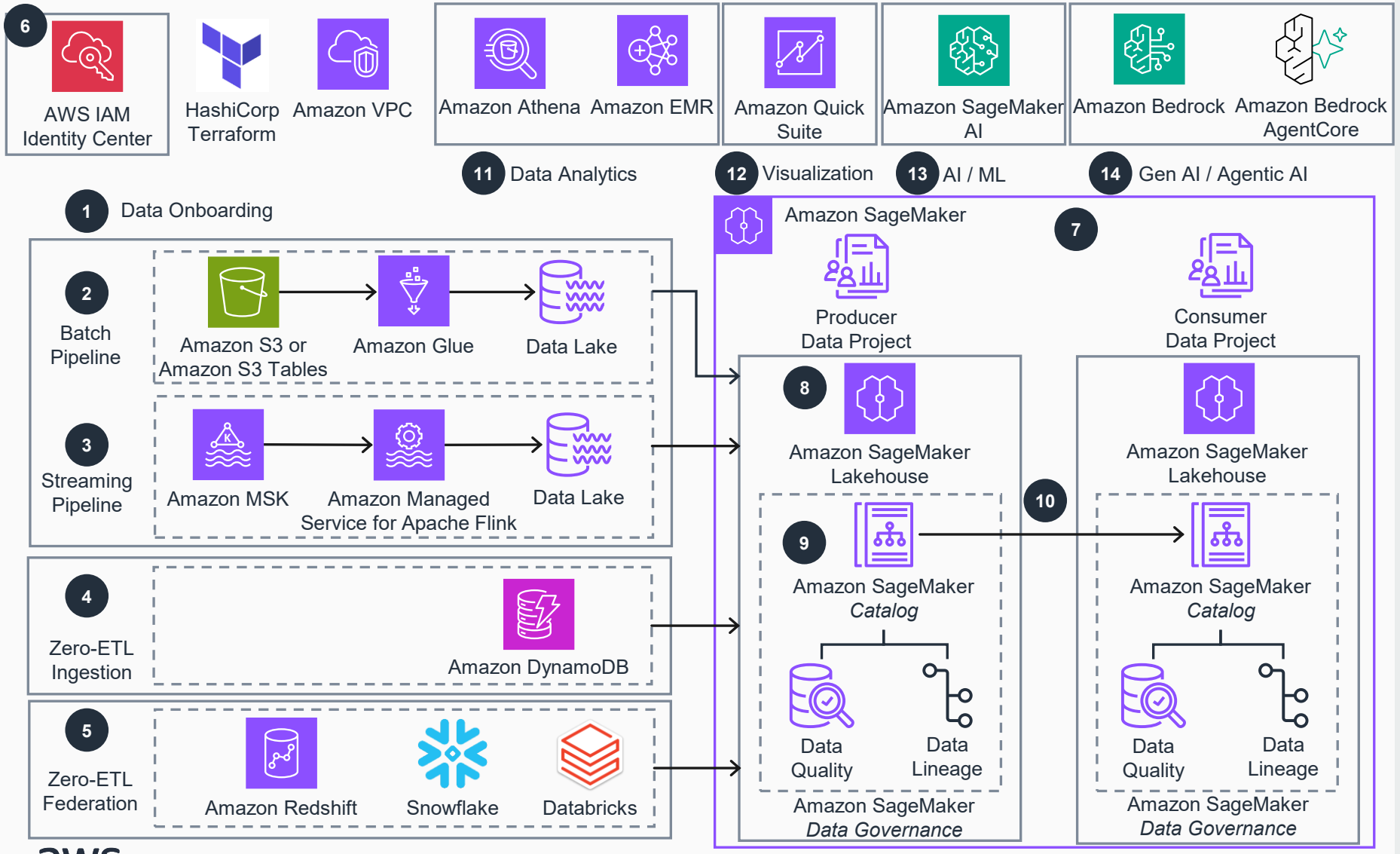
4 **Zero-ETL Ingestion:** Currently, Zero-ETL ingestion IaC modules include support for **Amazon DynamoDB**. Zero-ETL enables direct data movement between source and target data systems without the need to build and maintain ETL pipelines.

5 **Zero-ETL Federation:** Currently, Zero-ETL federation IaC modules include support for **Amazon Redshift**, Snowflake, and Databricks, enabling direct querying of data across these sources without data movement or replication, while maintaining consistent access controls and governance.

Reviewed for technical accuracy November 19, 2025
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Guidance for Developing a Data & AI Foundation with Amazon SageMaker

This reference architecture illustrates how to rapidly deploy data, analytics, AI, and visualization solutions on AWS, using Amazon SageMaker, Amazon Quick Suite, and Amazon Bedrock AgentCore. This solution provides the Infrastructure-as-Code (IaC) building blocks to simplify the process of building enterprise data platforms on AWS. This slide shows steps 6-11.



AWS Reference Architecture

6 **Identity Center:** This IaC module helps provision **AWS IAM Identity Center** instances at organization level or account level, create users and groups, and grant them required permissions. The user identities are needed login to **Amazon SageMaker** domain and Projects.

7 **Amazon SageMaker Unified Studio:** The **Amazon SageMaker** IaC module provisions **Amazon SageMaker** Domains and Projects, adds members to the projects, configures Lakehouse, and adds compute, with integration into **IAM Identity Center** to support various **Amazon SageMaker** domain or project roles. The solution creates starter Producer Data Project (to curate data assets), and Consumer Data Project (to consume data assets).

8 **Lakehouse**: Use **AWS Lake Formation** to federate data lakes into **Amazon SageMaker** Lakehouse or use zero-ETL integrations to make data available within the Lakehouse. **Amazon SageMaker** Lakehouse then provides a unified view across all data sources, where users can query data or build data products for downstream applications.

9 **Data Governance**: Use **Amazon SageMaker** Catalog to create curated data assets and to visualize data quality and data lineage. The governance policies within **Amazon SageMaker** are used to ensure data access and alignment with business requirements.

10 **Data Collaboration:** Use the producer project to produce data assets. Subscribe to data assets and consume them from the consumer project.

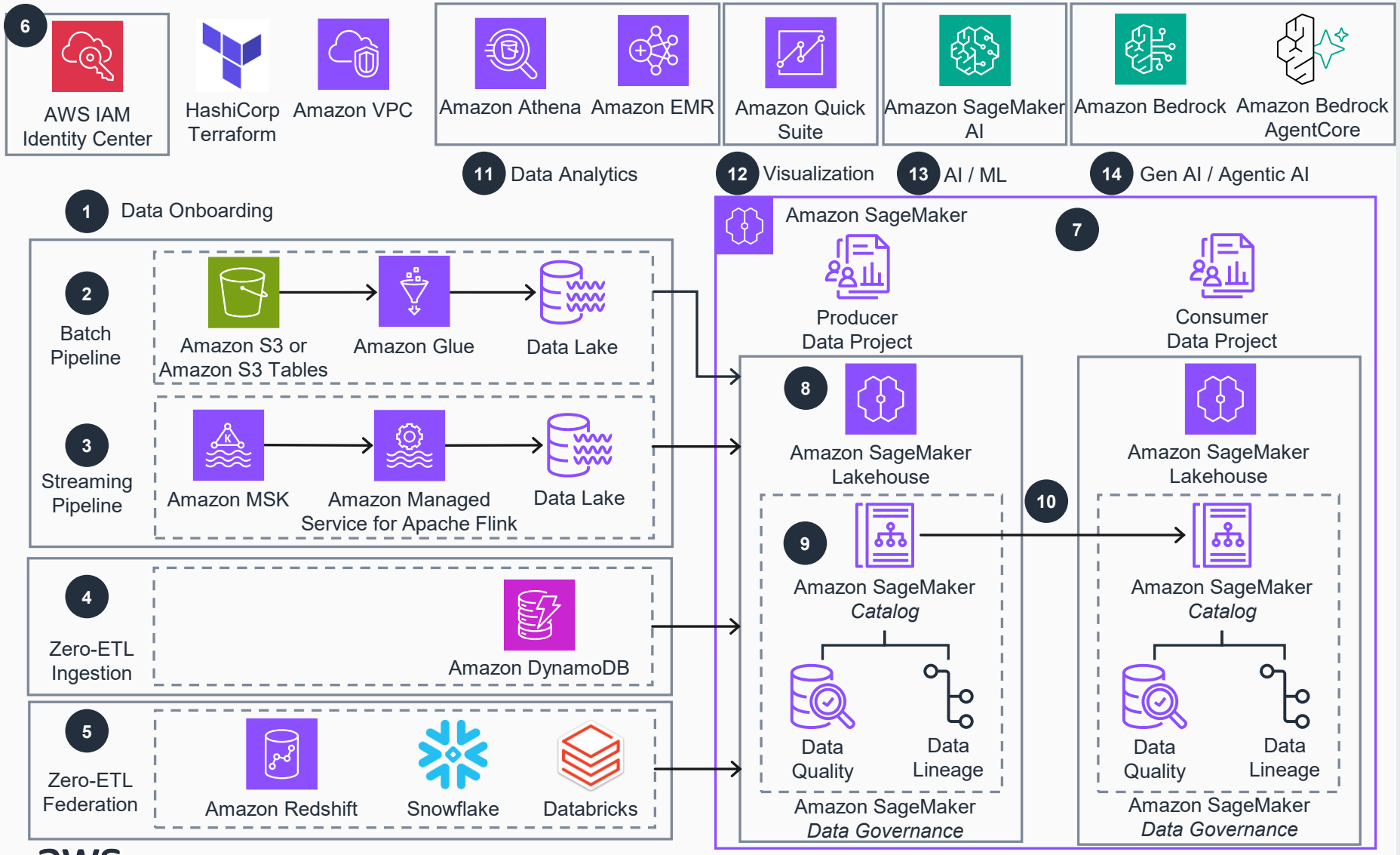
11 **Data Analytics:** Use **Amazon Athena** to query data from the data lake, hive, or iceberg, whether stored in an **Amazon S3** bucket or table. Implement deep analytics using **Amazon EMR**.

Reviewed for technical accuracy November 19, 2025

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

Guidance for Developing a Data & AI Foundation with Amazon SageMaker

This reference architecture illustrates how to rapidly deploy data, analytics, AI, and visualization solutions on AWS, using Amazon SageMaker, Amazon Quick Suite, and Amazon Bedrock AgentCore. This solution provides the Infrastructure-as-Code (IaC) building blocks to simplify the process of building enterprise data platforms on AWS. This slide shows steps 12-14.



AWS Reference Architecture

12 **Visualization:** Use **Amazon Quick Suite** to generate AI powered dashboards and reports using data in Lakehouse. **Amazon Quick Suite** provides unified intelligence across all your enterprise data sources and bridges the critical "last-mile gap" between insights and action. With the built-in agents for research and automation in **Amazon Quick Suite**, you can explore data conversationally and take actions directly from dashboards

13 **AI/ML:** Use **Amazon SageMaker AI** to develop Machine learning and AI applications.

14 **GenAI and Agentic AI: Amazon SageMaker** Unified Studio integrates with **Amazon Bedrock**, providing access to a range of high-performing foundation models (FMs) that can be used as the core intelligence for your agents. Leverage features like **Amazon Bedrock** Knowledge Bases, **Amazon Bedrock** Guardrails, **Amazon Bedrock AgentCore** (a comprehensive set of enterprise-grade services for securely deploying and operating AI agents at scale), and Flows within the Studio environment - to develop GenAI and Agentic AI applications – consuming data products from **Amazon SageMaker.**

Reviewed for technical accuracy November 19, 2025

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