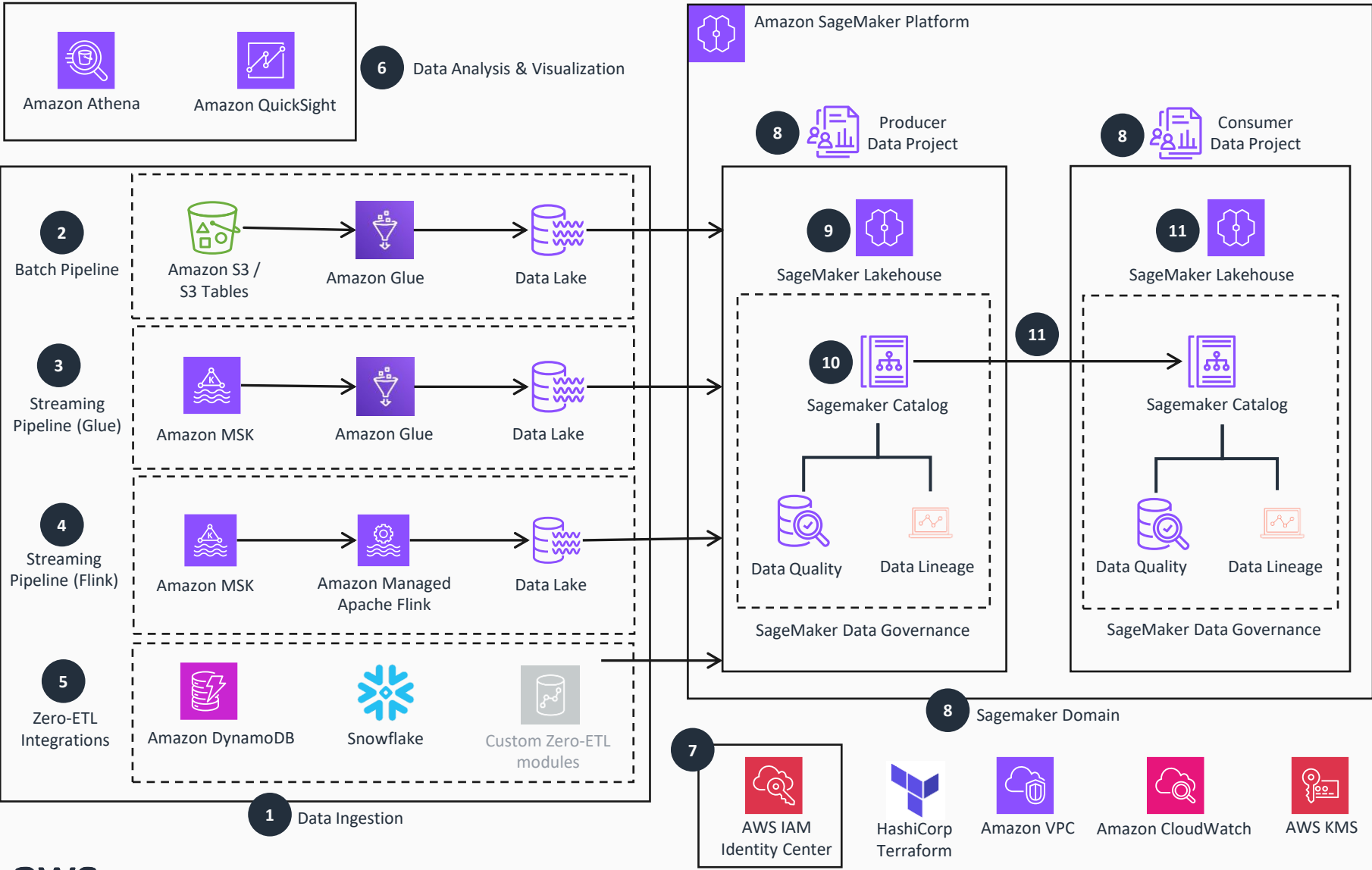


Guidance for Developing a Data & AI Foundation with Amazon SageMaker

This Guidance provides pre-built Terraform modules that help developers rapidly deploy data, analytics, AI, and visualization solutions on AWS, using Amazon SageMaker platform. This solution provides the Infrastructure-as-Code (IaC) building blocks to simplify the process of building enterprise data platforms on AWS.

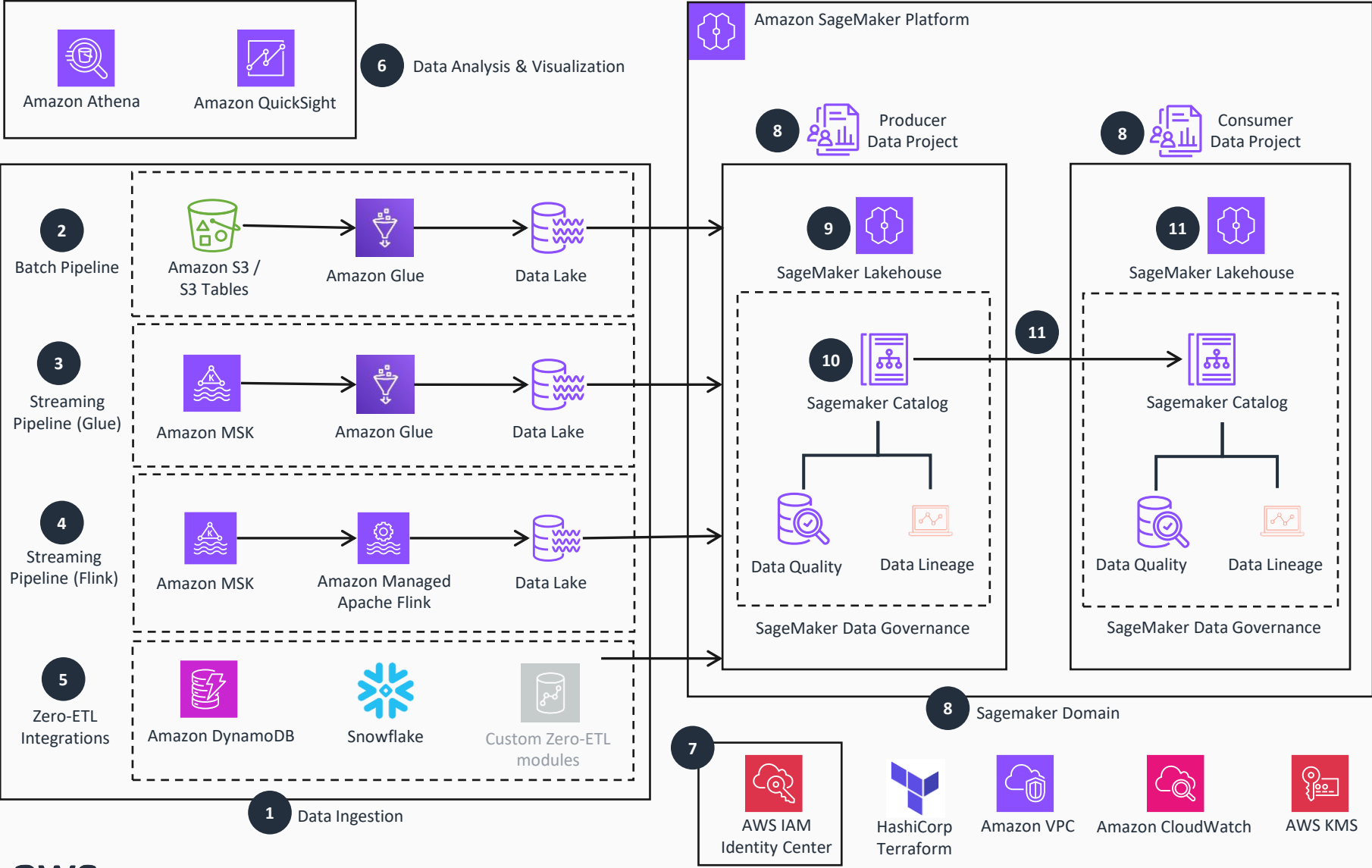


- 1 Data Ingestion:** This solution provides Infrastructure-as-Code (IaC) modules to provision data ingestion mechanisms. Customers can implement transactional Data Lakes, Schema evolution, and perform Time-travel queries on Iceberg data lakes, or setup Zero-ETL integrations.
- 2 Batch Pipelines:** Can be implemented using **AWS Glue** to ingest structured data from **Amazon S3 bucket** or **Amazon S3 table**. Support includes: Hive data lakes on standard S3 buckets and 3) Iceberg data lakes on S3 table buckets. Alternately, use EMR for batch data pipelines.
- 3 Streaming Pipeline (Glue):** Can be implemented using **AWS Glue Streaming** to ingest messages from **Amazon MSK** topics into hive or iceberg data lakes on S3 bucket or S3 table. Alternately, use **Amazon Data Firehose** to ingest messages from **Amazon MSK**.
- 4 Streaming Pipeline (Flink):** Can also be built using Apache Flink to ingest messages from MSK topics, perform streaming analytics, and store in hive or iceberg data lakes on **Amazon S3 bucket** or **S3 table**.
- 5 Zero-ETL Integrations:** Currently, Zero-ETL integration IaC modules includes support for DynamoDB and Snowflake, with more integrations coming. Zero-ETL enables direct data movement between source and target data systems without having to build and maintain ETL pipelines.
- 6 Data Analysis & Visualization:** Use **Amazon Athena** to query data from data lake, hive or iceberg, whether stored in S3 bucket or S3 table. Use **Amazon QuickSight** to generate reports and dashboards using data in data lakes.
- 7 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 Sagemaker domain and Projects.



Guidance for Developing a Data & AI Foundation with Amazon SageMaker

This Guidance provides pre-built Terraform modules that help developers rapidly deploy data, analytics, AI, and visualization solutions on AWS, using Amazon SageMaker platform. This solution provides the Infrastructure-as-Code (IaC) building blocks to simplify the process of building enterprise data platforms on AWS.



- 8 Sagemaker Unified Studio:** The **Amazon SageMaker** IaC module is used to provision Sagemaker Domains, Projects, add members to the projects, configure Lakehouse, add compute, with integration into IAM Identity Center to support various Sagemaker domain/project roles. The Unified Studio environment allows rich data, analytics and AI experience. The solution creates starter Producer Data Project (to curate data products), and Consumer Data Project (to consume data products).
- 9 Lakehouse:** Use **AWS Lake Formation** to federate data lakes into Sagemaker Lakehouse, or use Z-ETL integrations to make data available within the Lakehouse.
- 10 Governance:** Use SageMaker Catalog to create curated data products, visualize data quality and data lineage. The governance policies within SageMaker platform are used to ensure data access and alignment with business requirements.
- 11 Data Collaboration:** Use producer project to produce data products. Subscribe to data products and consume them from consumer lakehouse.

