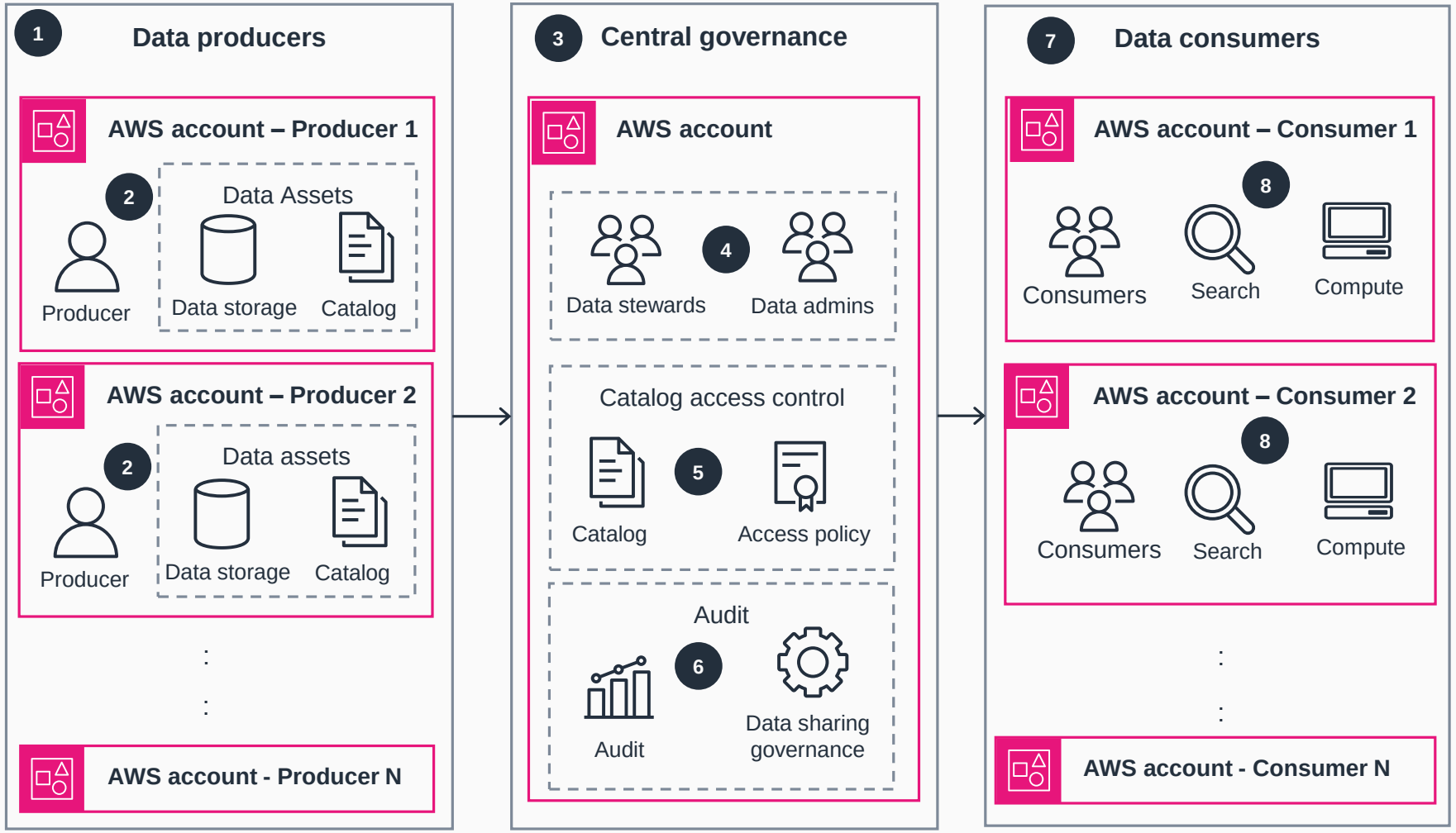
Guidance for a Secure Data Mesh with Distributed Data Asset Ownership on AWS

This architecture diagram illustrates an overview of a data mesh design that allows for distributed data ownership and control while providing centralized data sharing and governance to address security challenges. The subsequent diagram highlights the essential AWS services used in implementing this design pattern.



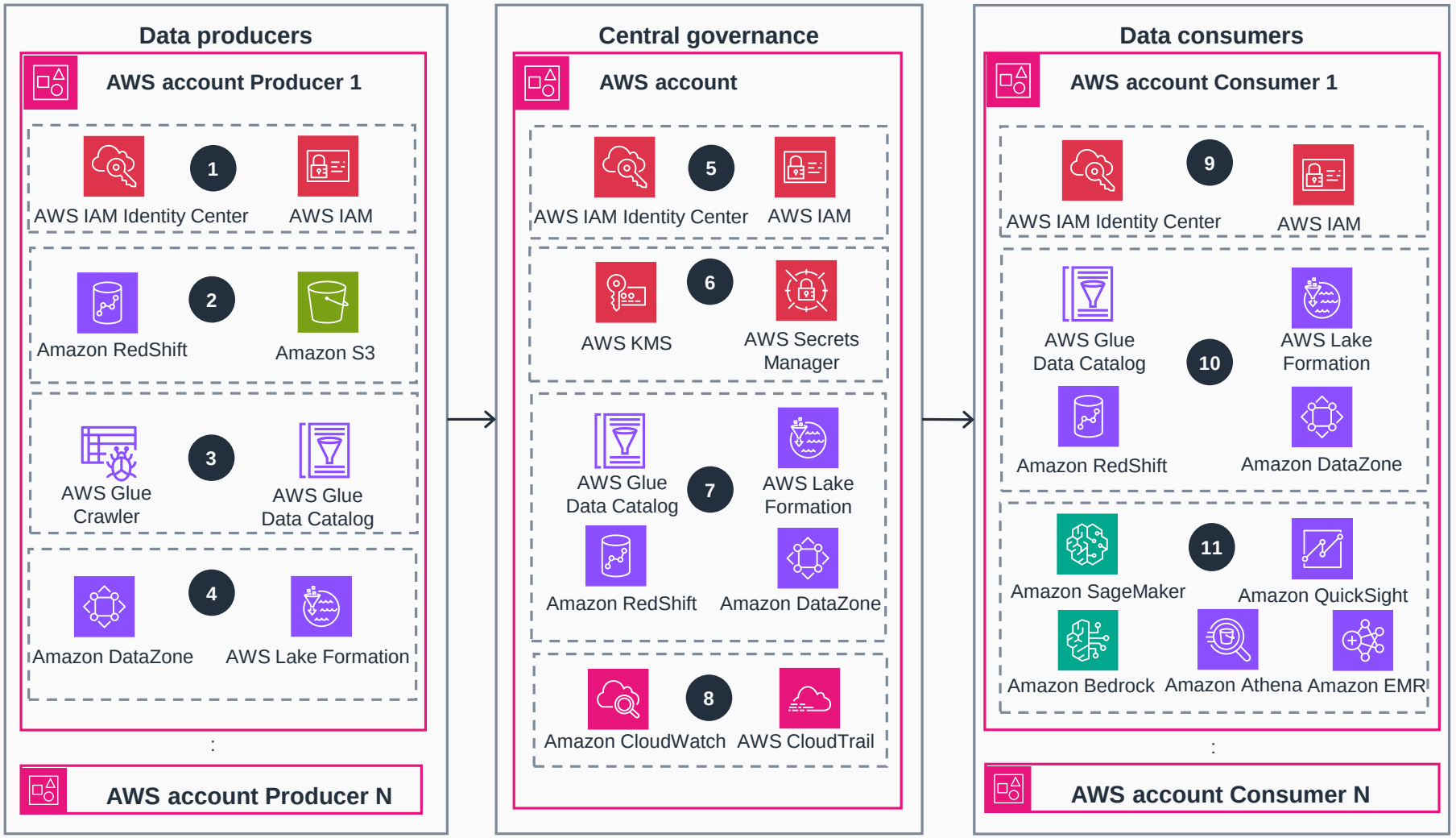
1. Multiple data producer accounts exist across different business domains and teams.
2. Data producers collect and transform data to generate shareable data assets, which mainly consist of a technical metadata catalog, databases, and scalable storage. It is the responsibility of the data producers to curate and keep the data assets current.
3. One central governance account acts as a bridge between data producers and data consumers. It does not save the actual data.
4. Data stewards maintain and enrich the enterprise data catalog across accounts with business metadata. Data admins create the necessary permissions for data producers to register data assets and data consumers to access data.
5. The central governance accounts maintain the enterprise data catalog and enrich the business catalog with corresponding access policies and encryption keys.
6. Central governance accounts save all the logs, including access logs and data shared object logs, and support audit reports.
7. Multiple data consumer accounts exist across different business domains and teams.
8. Data consumer accounts search the enterprise data catalog, request access to data assets, and bring their own compute resources to analyze the data once access is granted.

Reviewed for technical accuracy Month Day, 2024

AWS Reference Architecture

Guidance for a Secure Data Mesh with Distributed Data Asset Ownership on AWS

This architecture diagram shows the pivotal AWS services that allow the various components of this Guidance to function seamlessly within the data mesh architecture on AWS. This slide shows steps 1-8, refer to the next slide for steps 9-11.



1. Data producer users or roles authenticate through **AWS Identity and Access Management (IAM)** and/or single sign-on (SSO) providers like Okta and Azure Active Directory (Azure AD) integrated through **AWS IAM Identity Center**. Appropriate policies are attached to allow them to publish data assets.

2. Data assets that are ready to share are saved in scalable data storages like **Amazon Simple Storage Service (Amazon S3)** and **Amazon RedShift**.

3. Data producers use the **AWS Glue** crawler, which automatically generates technical metadata in the **AWS Glue** Data Catalog.

4. **Amazon DataZone** and **AWS Lake Formation** use the data catalog from **AWS Glue** and **Amazon Redshift** to generate shareable technical metadata.

5. Data stewards and data admins authenticate users and roles through **IAM** and/or SSO providers, which are integrated through the **IAM Identity Center**. Appropriate policies are attached to allow them to manage access.

6. **AWS Key Management Service (AWS KMS)** encrypts the data at rest and in transit. **AWS Secrets Manager** holds secrets like database credentials.

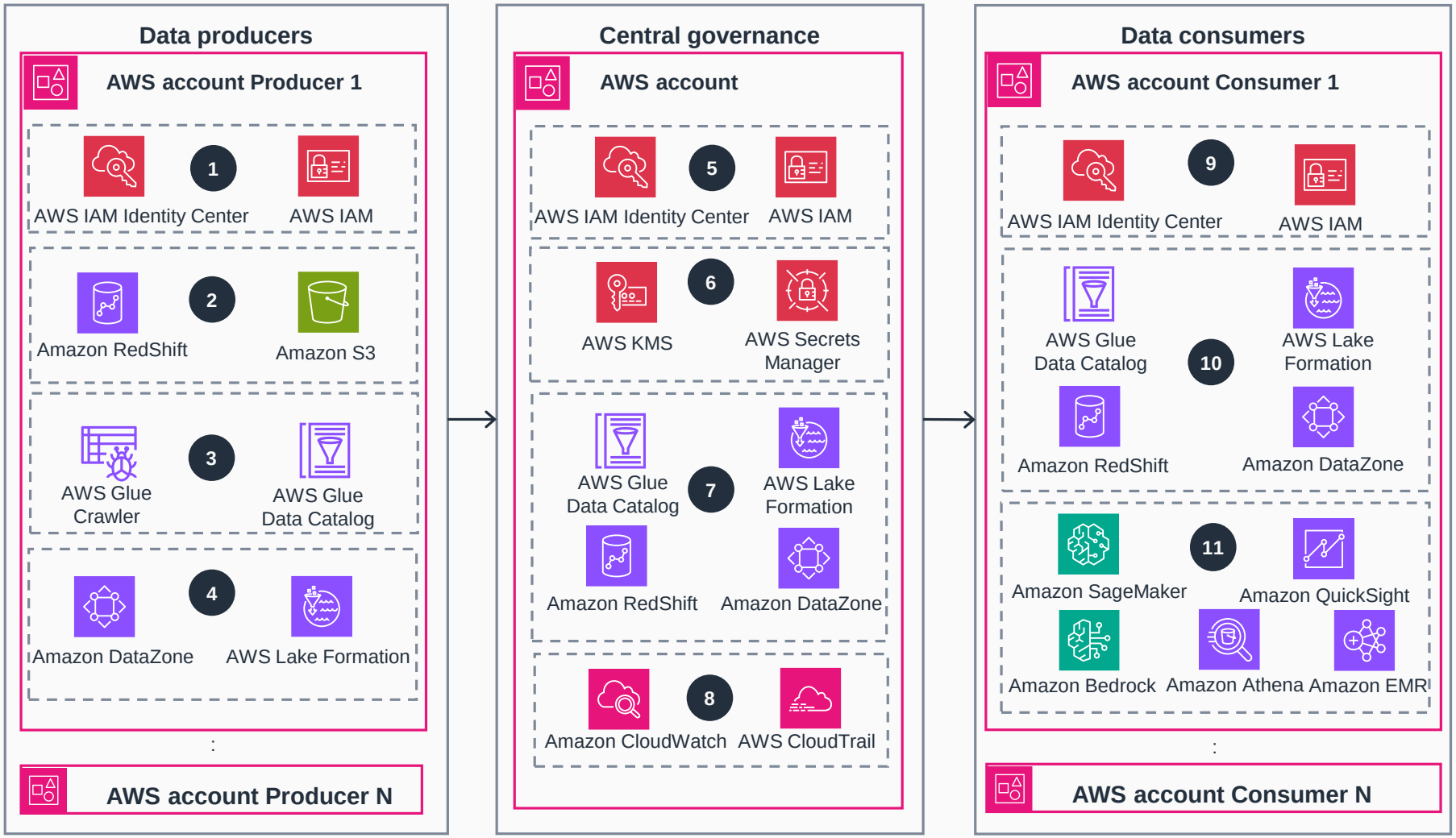
7. **Lake Formation** grants consumer users and roles access to producer data stored in **Amazon Redshift**. The **Amazon DataZone** domain enriches metadata stored in the **Data Catalog** by adding business metadata.

8. All of the access logs are available in **Lake Formation**, **Amazon CloudWatch**, and **AWS CloudTrail**, which users can utilize for monitoring and auditing.

AWS Reference Architecture

Guidance for a Secure Data Mesh with Distributed Data Asset Ownership on AWS

Steps 9-11.



9 **IAM** and/or SSO systems are integrated through **IAM Identity Center** to authenticate data consumer users and roles.

10 Consumers further granularize the access permissions using access permissions based on **Lake Formation**. Additionally, they use the **Amazon DataZone** domain to search for data assets based on metadata.

11 Consumers bring their own compute services. For instance, data scientists use **Amazon SageMaker** for machine learning (ML) transformation and **Amazon Bedrock** for generative artificial intelligence (AI) applications. Data engineers use **AWS Glue** and **Amazon EMR** for data transformation. Data analysts use **Amazon Athena** for analysis, and business intelligence analysts use **Amazon QuickSight** for data visualization.

Reviewed for technical accuracy Month Day, 2024

AWS Reference Architecture