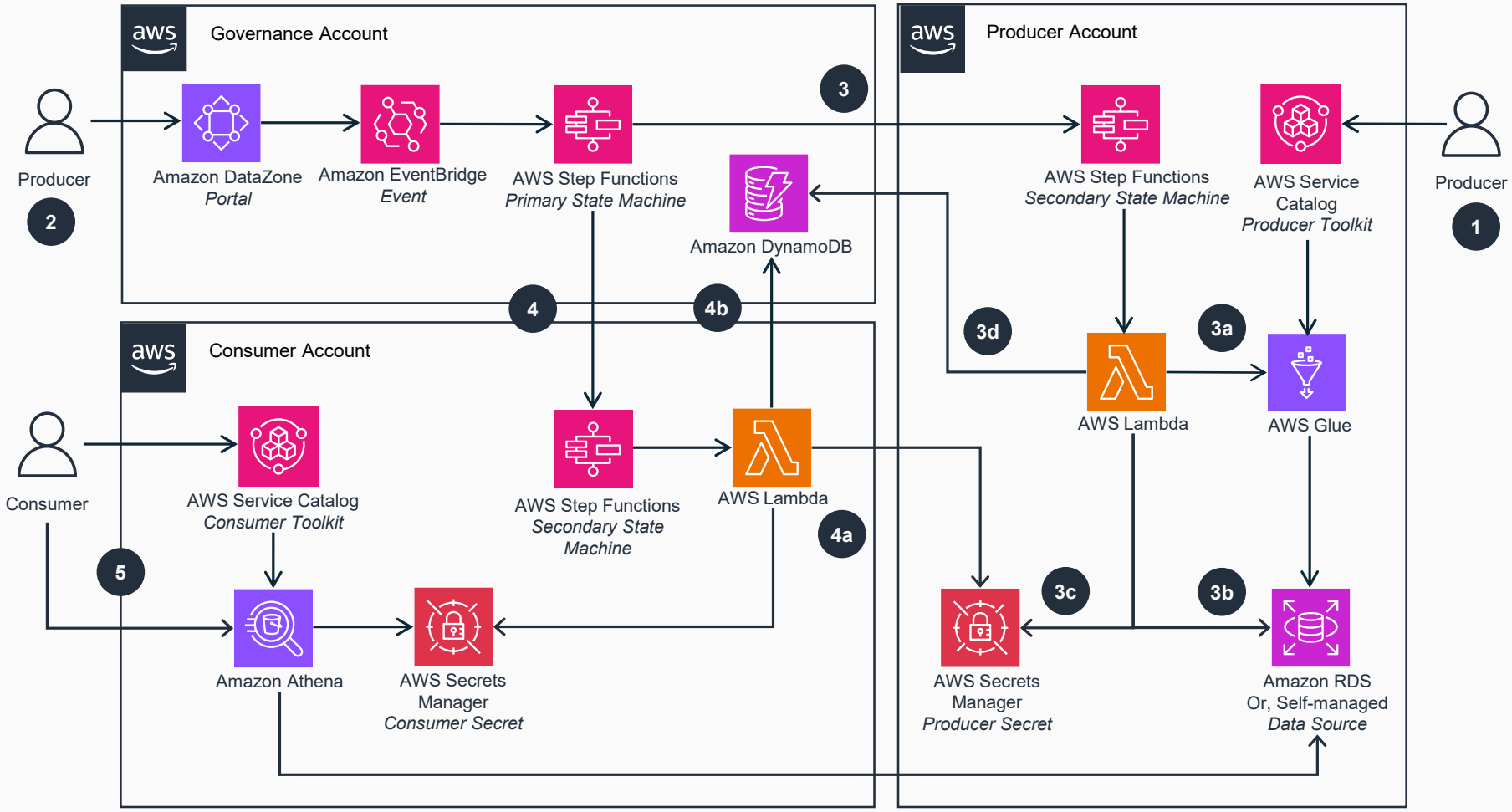
Guidance for Connecting Data Products with Amazon DataZone

This architecture diagram shows how you can extend the Amazon DataZone governance coverage to Java Database Connectivity (JDBC) data sources, including sources that are based on MySQL, PostgreSQL, Oracle, and SQL server engines. Steps 1-3 are displayed here. Continue to the next slide for steps 4-5.



1. A producer provisions a tool from the producer toolkit on **AWS Service Catalog** in the producer account. The tool will map data assets from the data source into the **AWS Glue** catalog.

2. The producer approves a subscription request for one of the mapped data assets in the **Amazon DataZone** portal. An event is sent to **Amazon EventBridge** and invokes an **AWS Step Functions** primary state machine in the governance account.

3. The primary state machine in the governance account invokes a **Step Functions** secondary state machine in the producer account.

3a. The secondary state machine in the producer account uses **AWS Lambda** to retrieve details for connecting to the data source hosting the subscription's data asset from **AWS Glue**.

3b. The secondary state machine in the producer account uses **Lambda** to connect to the data source, create credentials for the subscription's **Amazon DataZone** environment (if non-existent), and grant read access to the subscription's data asset.

3c. The secondary state machine in the producer account uses **Lambda** that persists the new data source credentials in an **AWS Secrets Manager** producer secret (if non-existent) with a resource policy allowing read and cross-account access to the **Amazon DataZone** environment's associated consumer account.

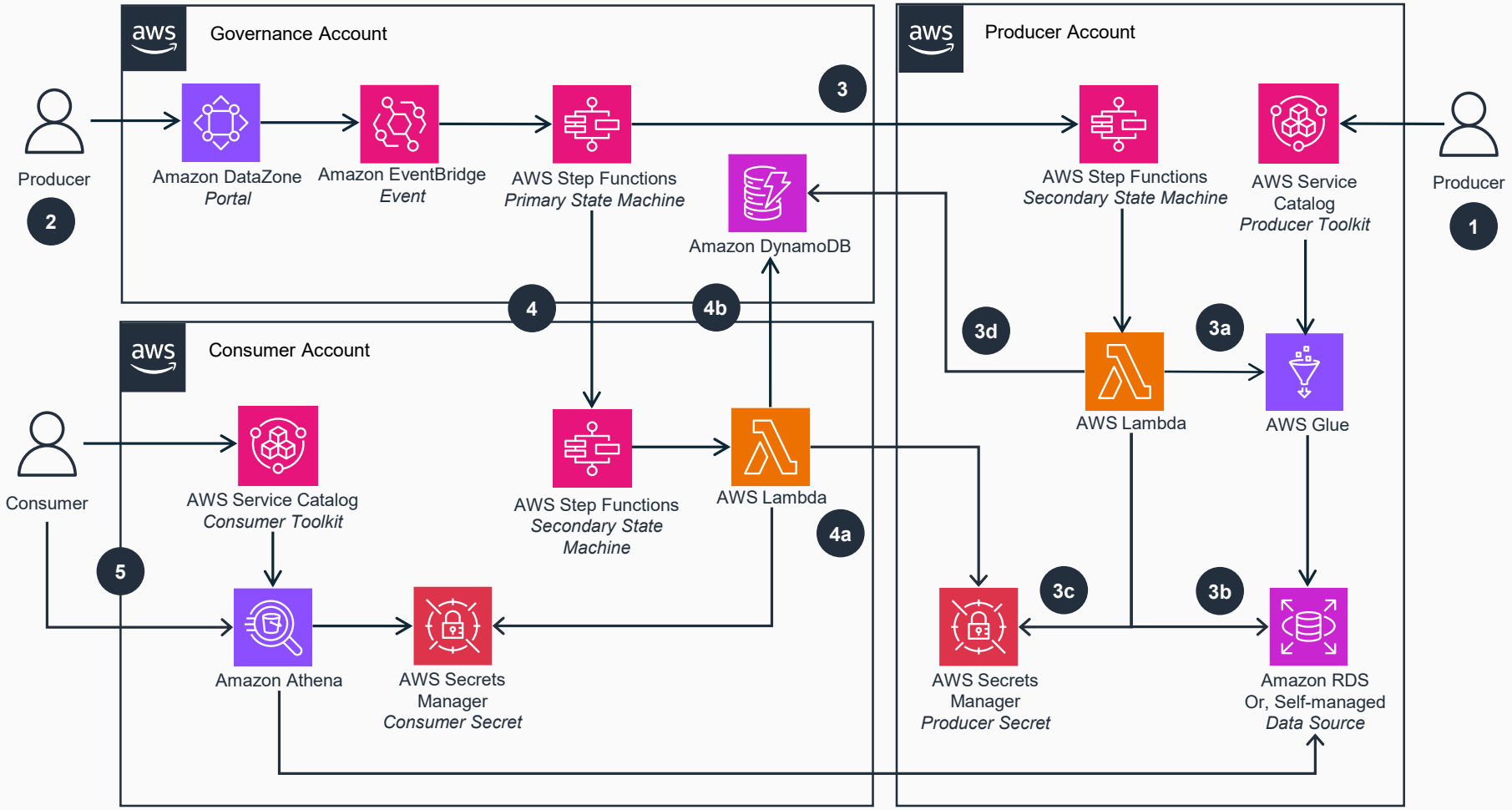
3d. The secondary state machine in the producer account uses **Lambda** to update tracking records on **Amazon DynamoDB** tables of the governance account.

Reviewed for technical accuracy October 18, 2023

AWS Reference Architecture

Guidance for Connecting Data Products with Amazon DataZone

Steps 4-5



4 The primary state machine in the governance account invokes a **Step Functions** secondary state machine in the consumer account.

4a The secondary state machine in the consumer account uses **Lambda** to retrieve connection credentials from the producer secret in the producer account through cross-account access. Then it copies the credentials into a new consumer secret (if non-existent) in **Secrets Manager** local to the consumer account.

4b The secondary state machine in the consumer account uses **Lambda** to update tracking records on **DynamoDB** tables in the governance account.

5 A consumer provisions a tool from the consumer toolkit in the consumer account on **Service Catalog**. The tool allows the consumer to query the subscription's data asset from its hosting data source through **Amazon Athena** by using the credentials stored in the consumer secret.

Reviewed for technical accuracy October 18, 2023

AWS Reference Architecture