

Guidance for CPG Data Products on AWS

This architecture uses a domain-owned design, giving CPGs a blueprint for organizing their data by domain and offering that data as a product to the rest of the enterprise. This approach helps CPGs accelerate time to actionable insights.

- 1 Collect data from multiple data sources and ingest this data into the different business domains across the enterprise.
- 2 Producer domains maintain ownership and autonomy of the data for their domain. Each domain creates metadata catalogs using **AWS Glue Data Catalog**. The producer domain controls data permissions.
- 3 Datasets from producer domains are presented as a product to the central governance account, and producers create **AWS Lake Formation** data catalog entities.
- 4 The central **AWS Lake Formation** data catalog shares the data catalog resources back to the producer account with required permissions through **AWS Lake Formation** resource links to metadata databases and tables.
- 5 Data changes made within the producer account are automatically propagated into the central governance copy of the catalog.
- 6 **Amazon Neptune** helps provide full context for the data through data lineage.
- 7 **Amazon DynamoDB** stores immutable data contracts. Consumers can subscribe to specific data contract attributes. **Amazon API Gateway** and **AWS Lambda** expose data to the enterprise as APIs.
- 8 Consumer domains can be the same as the producer domains or unique consumer domains. **Amazon Athena** and **Amazon QuickSight** query and visualize data. **Amazon Athena**, **Amazon EMR**, and **Amazon Redshift** perform advanced analytics. **Amazon SageMaker** helps build machine learning (ML) models.

