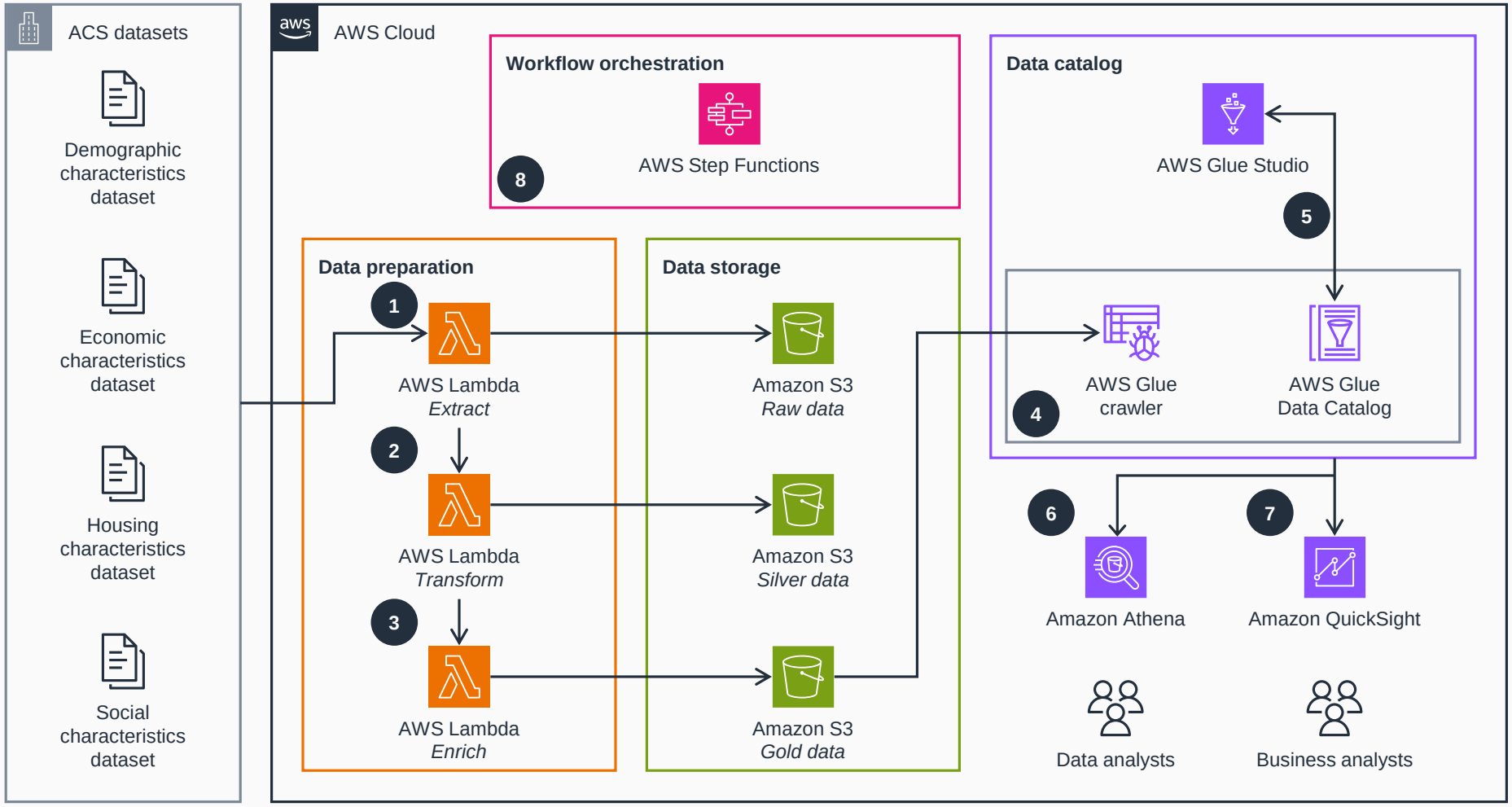
Guidance for Analyzing and Visualizing American Community Survey (ACS) Datasets

This architecture diagram shows how to extract, transform, enrich, store, and visualize ACS datasets (such as demographic, economic, housing, and social) using AWS data analytics services.



1. The **AWS Lambda** *Extract* function downloads and extracts ACS 5-year data using the US Census Bureau's API. Data is extracted for all US metropolitan areas and is stored in **Amazon Simple Storage Service (Amazon S3)**.
2. The **Lambda** *Transform* function removes rows and columns that are not needed for building visualizations, and it stores the data in an **S3** bucket. This dataset is called the silver data.
3. The **Lambda** *Enrich* function adds fields to the dataset, including postal codes and state information, then loads the data into an **S3** bucket. This dataset is called the gold data.
4. An **AWS Glue crawler** creates a schema for the gold data in the **AWS Glue Data Catalog**. Each ACS dataset is cataloged as a separate table in the catalog.
5. Use **AWS Glue Studio** to load, filter, transform, and save your partitioned data through SQL expressions or user-defined functions to further filter and transform data.
6. Data analysts can use **Amazon Athena** to run various one-time SQL queries on the ACS datasets cataloged in the **Data Catalog**. **Athena** provides interactive querying, analyzing, and processing capabilities.
7. Business analysts can use **Amazon QuickSight** to build machine learning (ML)–powered business intelligence visualizations and dashboards showing key demographic, economic, housing, and social metrics about the communities they serve.
8. **AWS Step Functions** orchestrates the AWS services that prepare, store, and catalog data.

Reviewed for technical accuracy March 13, 2024

AWS Reference Architecture