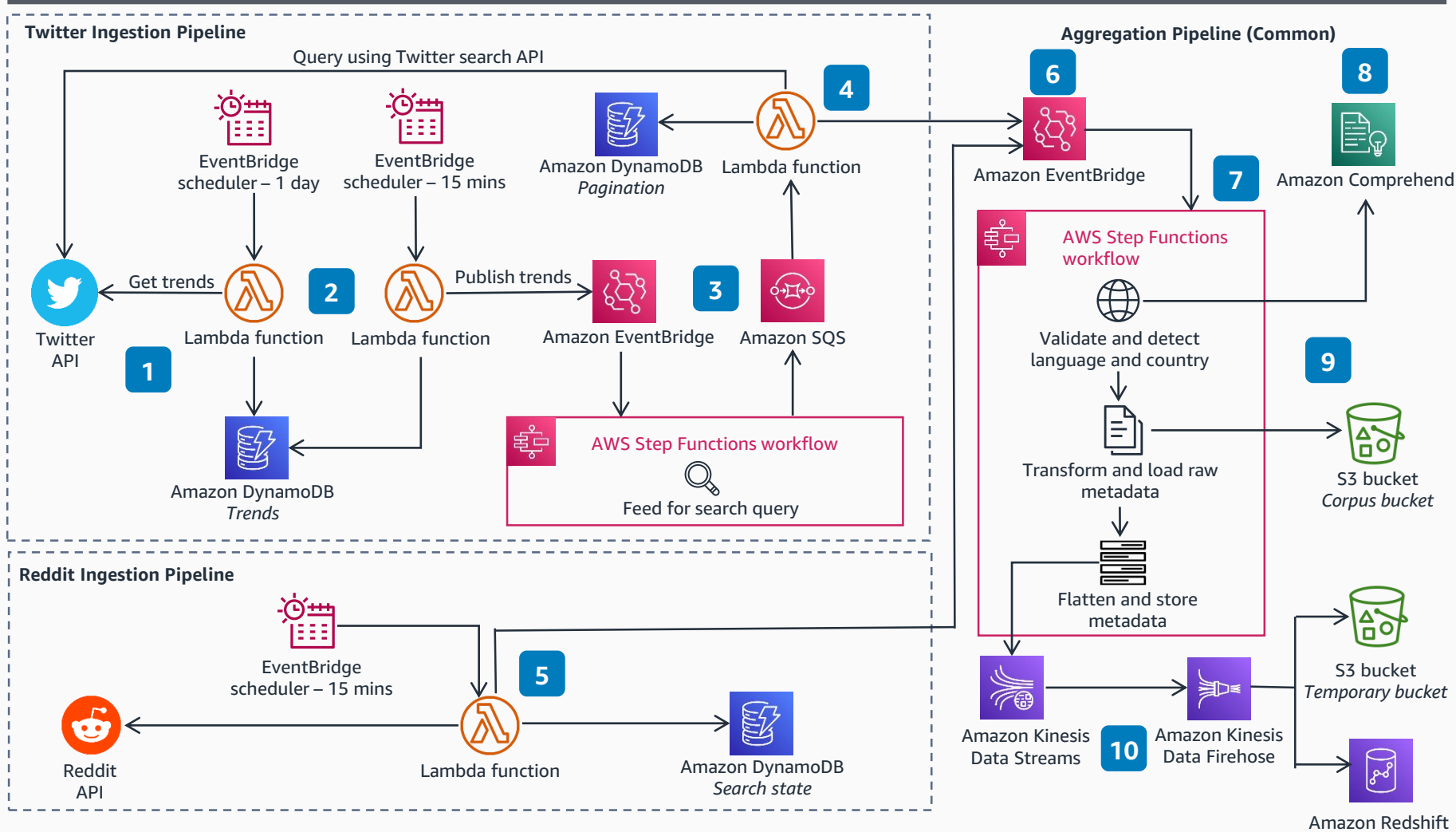


# Guidance for Social Media Data Pipeline on AWS

This modular architecture uses AWS serverless services to help you integrate with social media platforms and derive data insights for improved business outcomes.



- 1** Amazon EventBridge rules activate AWS Lambda functions to get the Twitter trends for a specific region, which are stored in Amazon DynamoDB. AWS Systems Manager (parameter store) is used for inputs such as location and credentials.
- 2** Scheduled Lambda functions get the latest trends from DynamoDB and publish them as events to EventBridge event bus for processing. EventBridge supports loosely-coupled event-driven applications at scale.
- 3** EventBridge rules activate AWS Step Functions for each event. Step Functions validates, tracks, and pushes the message to Amazon Simple Queue Service (Amazon SQS) to retrieve tweets. SQS is used in this architecture to manage the Twitter API throttling and errors.
- 4** For the batch of SQS messages, Lambda fetches the tweets using Search API and transforms and pushes the messages to EventBridge. DynamoDB stores "pagination" tokens and other details.
- 5** EventBridge invokes Lambda to fetch and transform popular subreddits and comments and publish them to EventBridge. DynamoDB stores search state.
- 6** EventBridge receives the messages from all sources in a common format.
- 7** Each message initiates Step Functions to validate, transform, and load the data to the targeted storage location.
- 8** Amazon Comprehend detects the language model if not present in the message.
- 9** Amazon Simple Storage Service (Amazon S3) provides durable and scalable storage for corpus.
- 10** As an optional step, use Amazon Kinesis to push the metadata information to other sources, such as Amazon Redshift, for better analytics and querying.