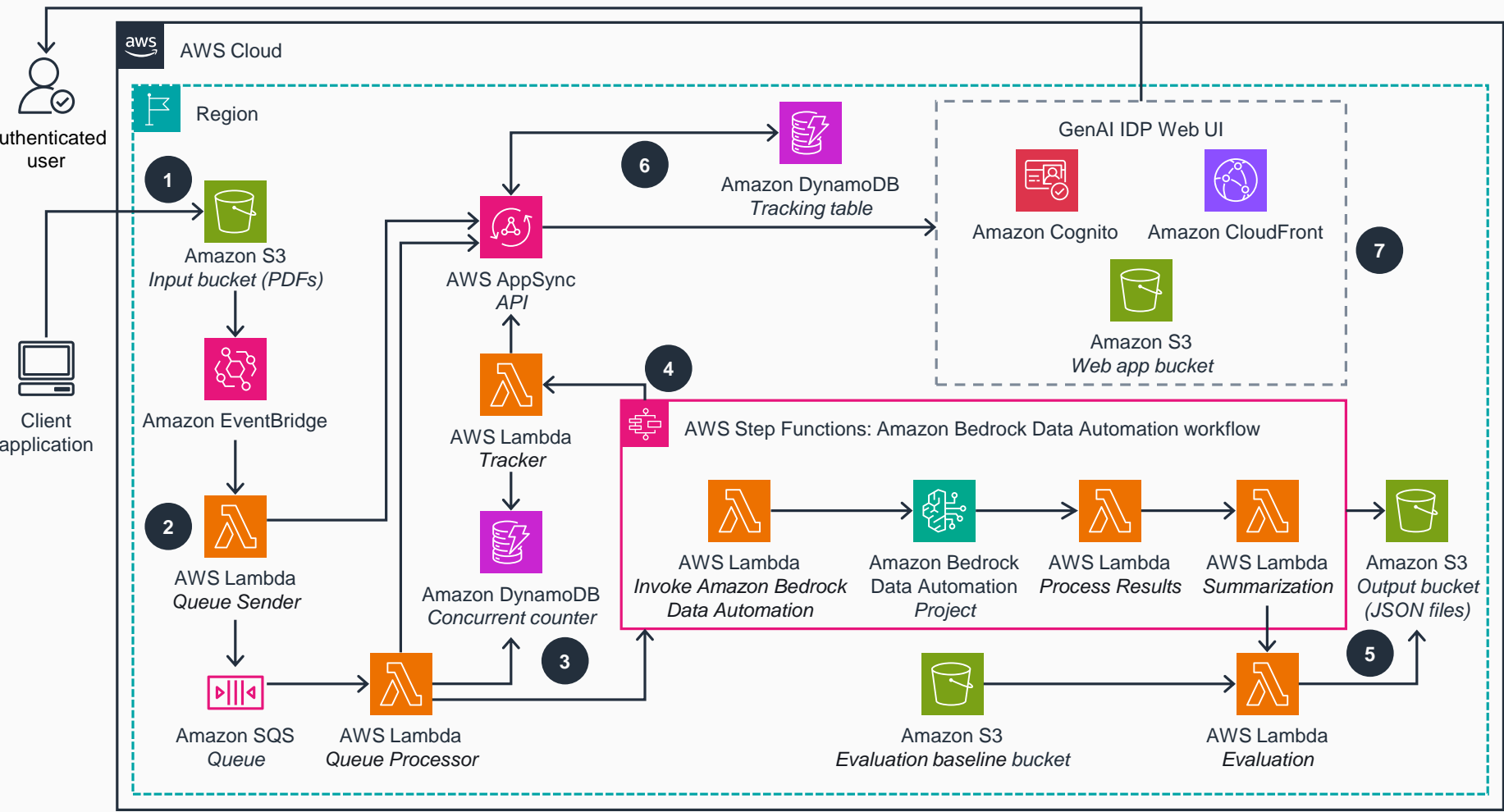# Guidance for Accelerated Intelligent Document Processing on AWS

This architecture diagram shows how to build a scalable document processing system that uses flexible generative AI model patterns to extract structured data from PDFs while processing data, tracking real-time status, and evaluating accuracy.

1. Client applications submit PDF documents through either the web interface or directly to the **Amazon Simple Storage Service (Amazon S3)** input bucket. **Amazon EventBridge** detects these uploads and invokes the document processing workflow.

2. **AWS Lambda** (Queue-Sender function) records document events in the **AWS AppSync** API for tracking and sends them to an **Amazon Simple Queue Service (Amazon SQS)** queue for message processing.

3. The **Lambda** (Queue-Processor function) retrieves messages from the **Amazon SQS** queue in batches. It manages workflow concurrency using an **Amazon DynamoDB** counter and initiates **AWS Step Functions** implementations for document processing.

4. The **Step Functions** workflow orchestrates document processing using **Lambda** functions. For this pattern, a **Lambda** function invokes **Amazon Bedrock Data Automation** processing to perform AI-powered document analysis tasks. Then, **Lambda** (Process-Results function) handles the output.

5. Upon workflow completion, **Lambda** (Evaluation function) automatically compares processing outputs against predefined baseline documents. These are known correct outputs in the **Amazon S3** evaluation-baseline bucket. Final results and evaluation reports are stored as JSON files in the **Amazon S3** output bucket.

6. The **AWS AppSync** API maintains document status in the **DynamoDB** tracking table and enables real-time status updates through the web interface. The tracker function updates processing metrics and status.

7. The web interface (GenAI IDP Web UI) is hosted in the **Amazon S3** web-app bucket and distributed through **Amazon CloudFront**. It uses **Amazon Cognito** for user authentication, enabling you to monitor document processing through a secure dashboard.

AWS Reference Architecture