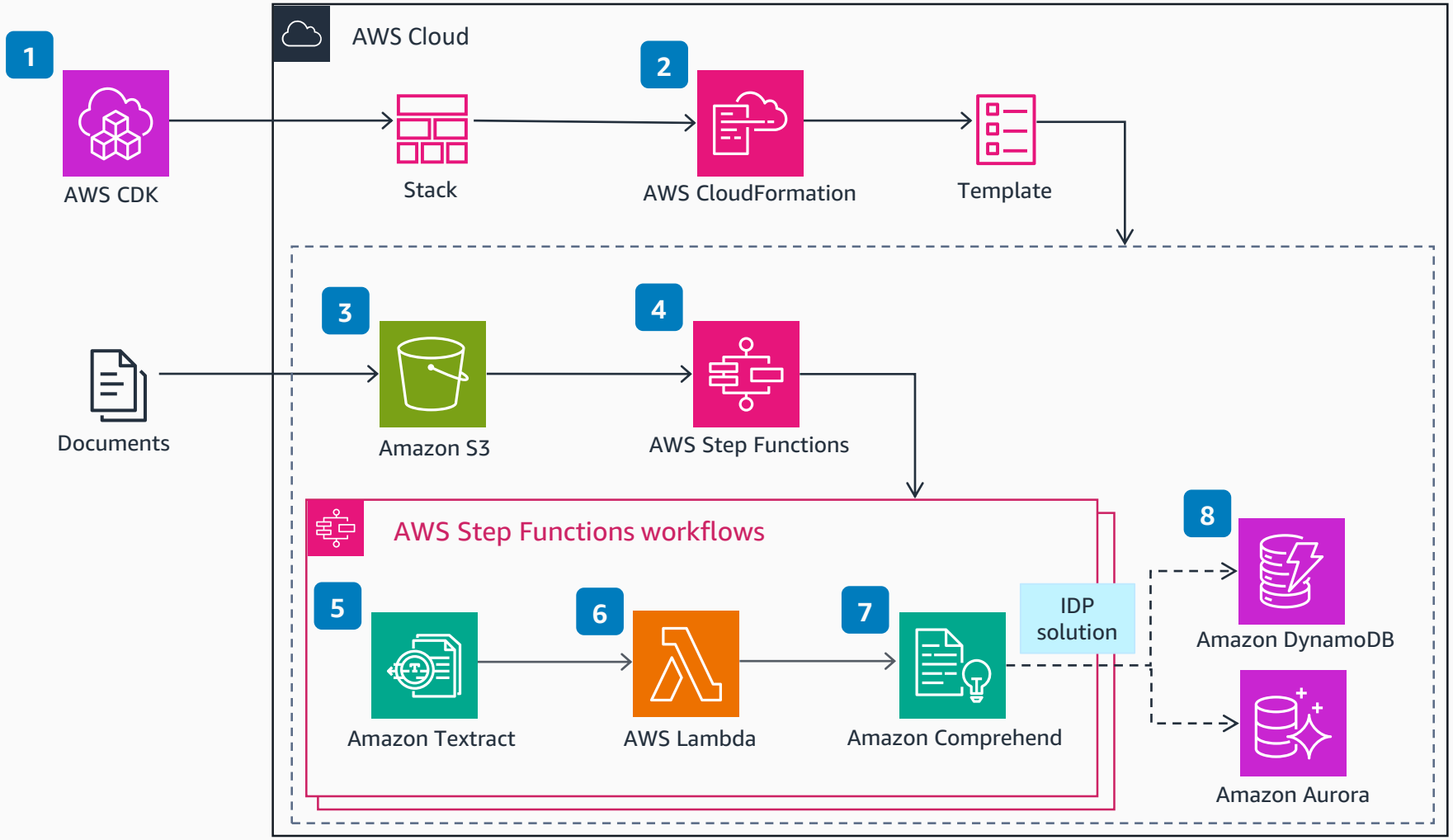
Guidance for Low-Code Intelligent Document Processing on AWS

This self-contained and robust architecture uses ML best practices to automate and process documents at scale.



1. Document processing workflows are developed using the **AWS Cloud Development Kit (AWS CDK)**.
2. **AWS CDK** generates a stack in **AWS CloudFormation** that deploys templates for the resources required to execute the workflows.
3. The **CloudFormation** template creates an **Amazon Simple Storage Service (Amazon S3)** bucket where documents are uploaded for processing.
4. Document uploads trigger an **AWS Step Functions** workflow to orchestrate document processing functions. Depending on your specific use case, you can choose one or more workflows available in the sample code.
5. The **Step Functions** workflow begins with **Amazon Textract** extracting the text from the document.
6. An **AWS Lambda** function processes the output of **Amazon Textract** and generates a CSV file and key/value pairs of extracted text.
7. **Amazon Comprehend** uses the extracted text to classify documents by type. The sample code includes custom classifiers.
8. Depending on the deployed workflow, **Amazon DynamoDB** or **Amazon Aurora** persists the data.

Reviewed for technical accuracy May 30, 2023

AWS Reference Architecture