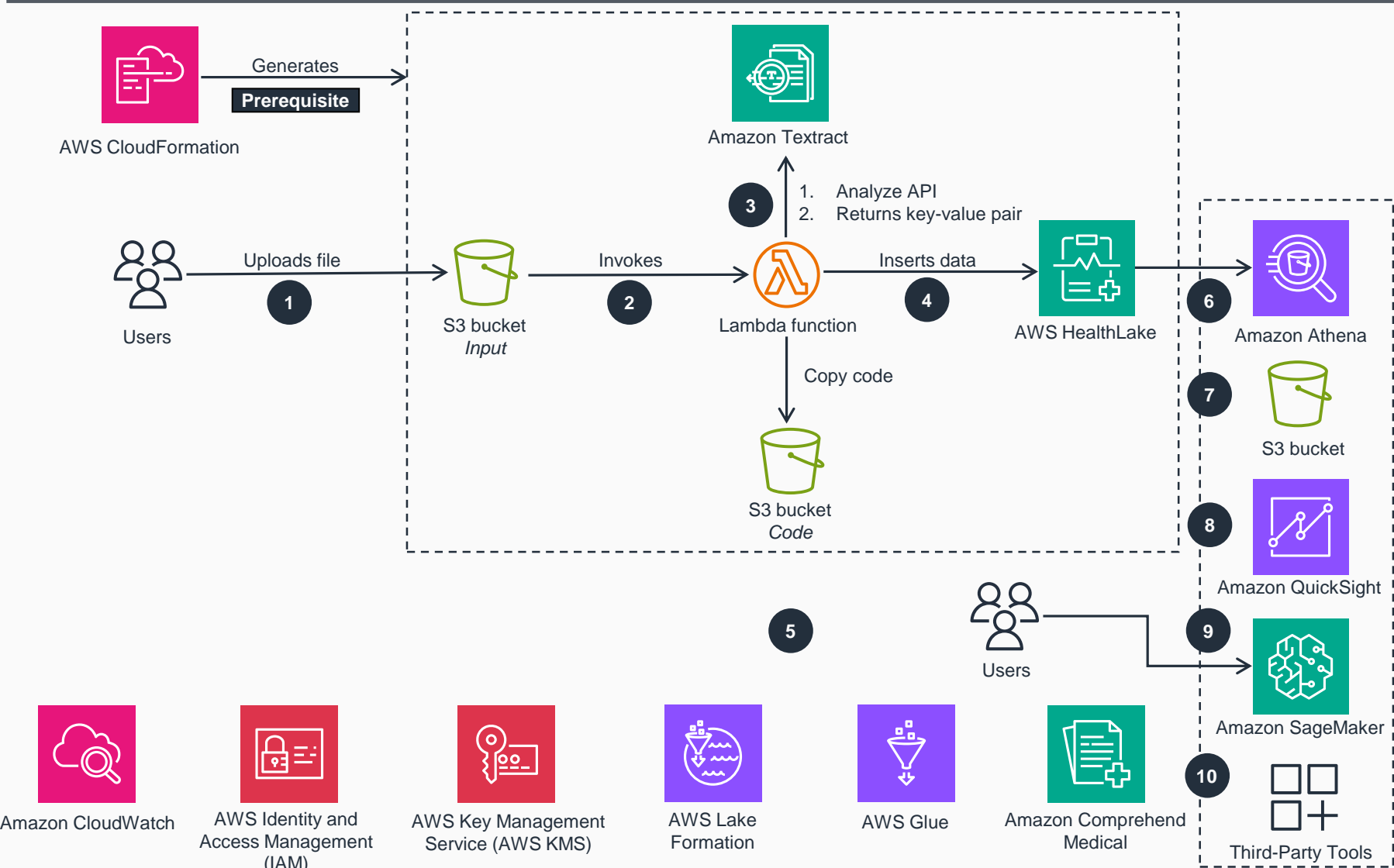


Guidance for Ingesting PDF and Image Files to AWS HealthLake

This architecture diagram shows how PDF and image files get ingested to AWS HealthLake for more efficient analysis.



Prerequisite

AWS CloudFormation generates various AWS resources.

- 1 Users upload a PDF file or an image file, ending with .tiff, .jpg, or .png.
- 2 Creating a new file in **Amazon Simple Storage Service (Amazon S3)** invokes an **AWS Lambda** function.
- 3 The **Lambda** function calls **Amazon Textract** APIs to parse the data.
- 4 The **Lambda** function inserts the key-value pairs into **AWS HealthLake**.
- 5 **HealthLake** transforms customer data into an analytics-ready format for **AWS Lake Formation** in near real-time to provide enhanced analytics capabilities. **Lake Formation** is built on **AWS Glue**. **HealthLake** automatically provides you with integrated natural language processing (NLP) using **Amazon Comprehend Medical**.
- 6 **Amazon Athena** queries the **HealthLake** data.
- 7 An **S3** bucket stores the **HealthLake** data.
- 8 Users can view the **HealthLake** data in a dashboard using **Amazon QuickSight**.
- 9 Users can build, train, and deploy machine learning (ML) models with **Amazon SageMaker** using **HealthLake** data.
- 10 **HealthLake** integrates with third-party tools.

