Workflow of the Application

Uploading Documents

1. Amplify (Frontend App)
- The user begins by logging into the web application hosted on Amplify.
- Amplify provides a globally available interface, abstracting away infrastructure concerns.
- From the UI, the user clicks Upload and selects a file.

2. Cognito (Authentication)
- Before the upload, the user is authenticated using Cognito (Authorization Code Flow with PKCE).
- This ensures only authorized users can access upload functionality, preventing anonymous access to storage.

3. API Gateway (Secure Entry Point)
- The frontend calls an API route, e.g., /uploads/create.
- API Gateway acts as a front door, verifying the Cognito JWT token before passing the request downstream.
- This avoids exposing S3 directly to the internet.

4. IAM (Permissions Management)
- IAM roles restrict what each service can do.
- The Lambda function handling uploads only has permission to generate a presigned URL for S3 and write to DynamoDB.
- This least-privilege setup protects against abuse or escalation.

5. Lambda (Upload Orchestration)
- Lambda validates the request and generates a presigned S3 URL.
- The presigned URL allows the user to upload directly to S3, avoiding any bottleneck in your API.
- Lambda may also insert an initial record into DynamoDB to track the file status as pending.

6. S3 Bucket (File Storage)
- The file is uploaded straight from the browser to the S3 bucket using the presigned URL.
- S3 stores the file securely (encryption at rest, TLS-only access) and emits an Object Created event.

7. EventBridge (Event Router)
- The S3 event is forwarded to EventBridge, which acts as the central event bus.
- EventBridge rules determine which processing workflow to trigger.

8. Step Functions (Processing Orchestration)
- Step Functions coordinate the metadata extraction steps.
- Example workflow:
- Validate the file.
- Extract metadata (size, type, version).
- Run specialized extractors for PDFs, images, or CSVs.
- Aggregate results.
- Step Functions provide built-in retries and error handling, ensuring reliability.

9. Lambda (Processing Tasks)
- Within Step Functions, Lambdas perform specific jobs (e.g., read EXIF tags, parse CSV headers, count PDF pages).
- These functions remain small and idempotent, so they can be retried safely.

10. DynamoDB (System of Record)
- Final metadata is written into DynamoDB.
- Each file has a record containing: project ID, filename, content type, size, status, and extracted attributes.
- DynamoDB provides fast lookups for your app's "Your Documents" page.

Displaying Documents

1. Amplify (Frontend)
- The user navigates to the "Your Documents" page.

2. API Gateway
- The frontend calls a secure endpoint (e.g., /files/list) with the user's Cognito token.

3. Lambda (List Files)
- The Lambda queries DynamoDB using projectId and pagination keys.
- It returns a lightweight list of metadata objects.

4. DynamoDB
- DynamoDB responds quickly, even at scale, enabling smooth UI updates.
- The app displays the file list with names, status, and metadata.

Downloading Documents

1. Amplify (Frontend)
- The user selects a file to download.

2. API Gateway
- The frontend calls a secure endpoint (e.g., /downloads/create).

3. Lambda (Download URL Generator)
- The Lambda looks up the file in DynamoDB to confirm existence and permissions.
- It generates a presigned S3 GET URL with a short expiration.

4. S3 Bucket (Data Transfer)
- The browser uses the presigned URL to download the file directly from S3.
- This avoids routing large files through the API, saving cost and improving performance.

Why This Workflow Matters

- Security: Cognito + API Gateway + IAM enforce least-privilege and prevent public S3 access.
- Scalability: S3, DynamoDB, and Lambda scale automatically without server maintenance.
- Reliability: Step Functions provide retries and visibility into processing.
- Performance: Large file transfers bypass the API, flowing directly between browser and S3.
- Extensibility: Adding support for new file types requires only new Lambda tasks and Step Functions branches.