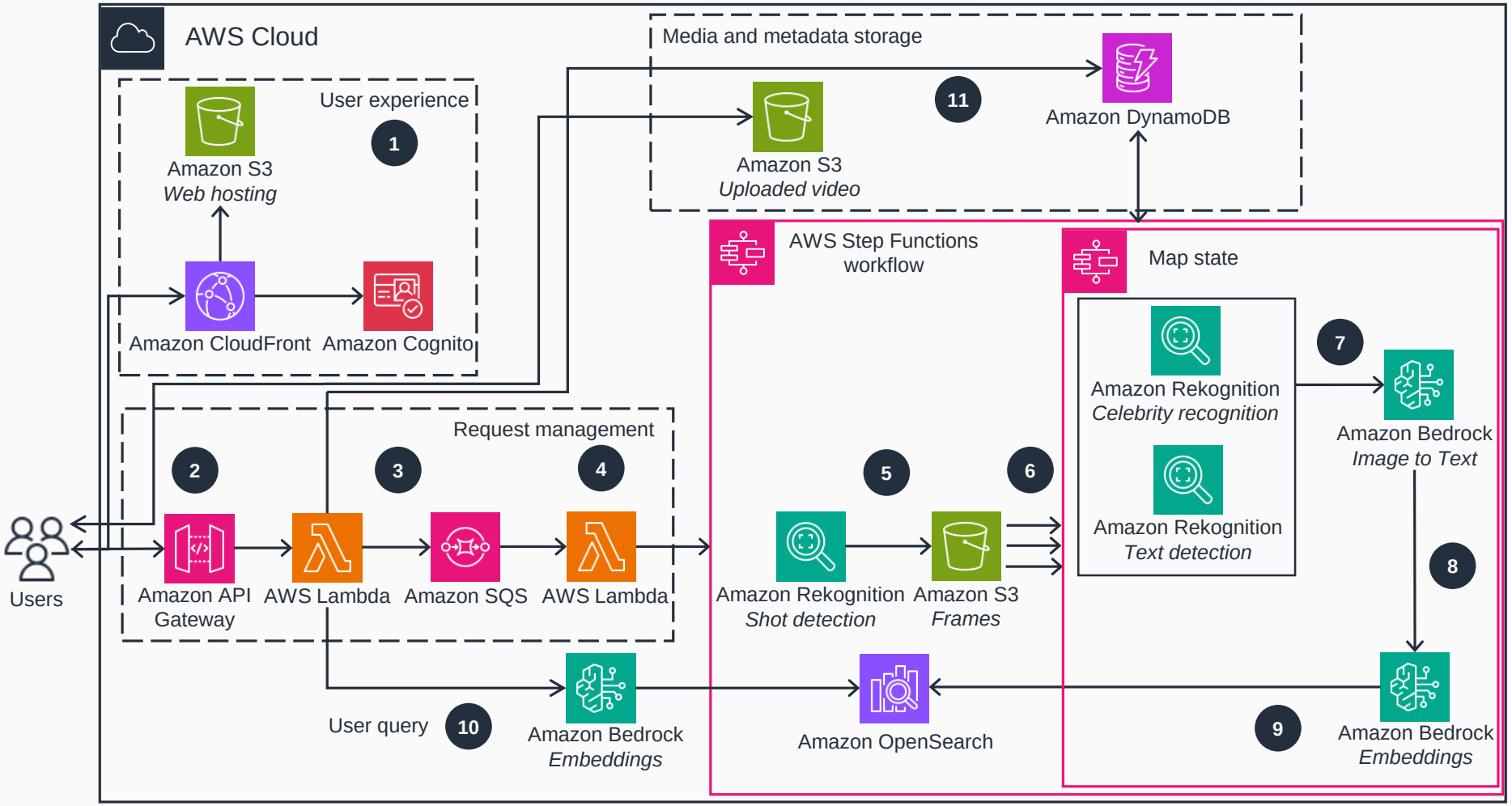
Guidance for Semantic Video Search on AWS

This architecture diagram shows a semantic video search workflow that allows users to quickly search for particular scenes, actions, people, objects, and more in video data using natural language queries. It uses AWS artificial intelligence and machine learning (AI/ML) services to process videos into a segmented array of frames. Users can then identify text and celebrities within each frame before using foundation models (FMs) to understand the context and meaning of the video segments, which are then embedded into vector search functionality. Steps 1-7 are shown here. For steps 8-11, go to the next slide.



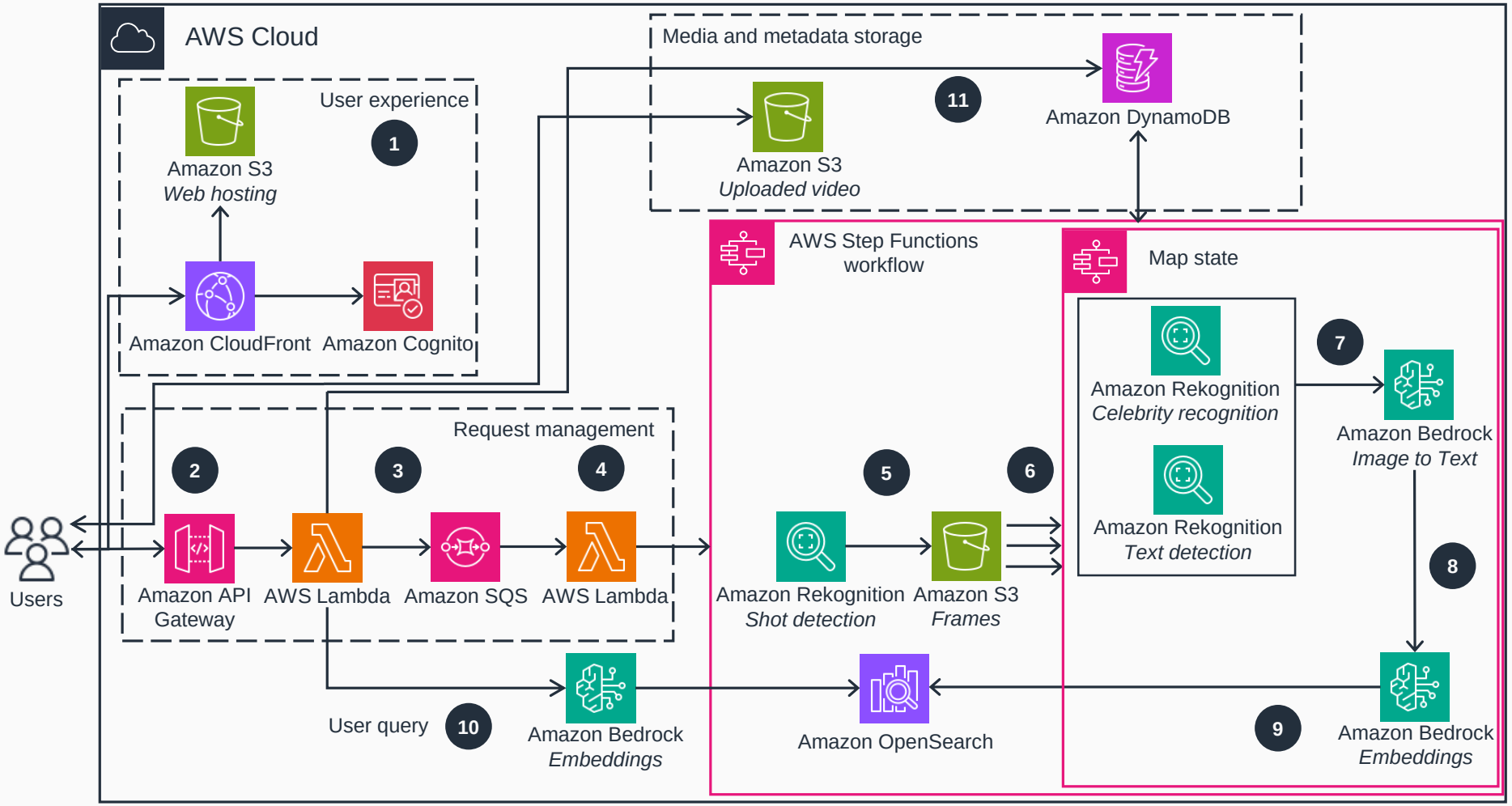
1. **Amazon Simple Storage Service (Amazon S3)** hosts a static website for the semantic video search, served by an **Amazon CloudFront** distribution. **Amazon Cognito** provides customer identity and access management for the web application.
2. Upload videos to **Amazon S3** with **Amazon S3** pre-signed URLs.
3. After a video is uploaded successfully, an API call to **Amazon API Gateway** invokes **AWS Lambda** to queue new indexing-video requests in **Amazon Simple Queue Service (Amazon SQS).**
4. **Lambda** processes new messages in the **Amazon SQS** queue, initiating an **AWS Step Functions** workflow.
5. **Amazon Rekognition** detects multiple video shots from the original video, containing the start, end, and duration of each shot. Shot metadata is used to generate a sequence of frames, which are grouped by individual video shots and stored in **Amazon S3**.
6. **Step Functions** uses the Map state to run a set of workflows for each frame stored in **Amazon S3** in parallel.
7. **Amazon Rekognition** detects celebrities and text in the frames.

Reviewed for technical accuracy April 23, 2024

AWS Reference Architecture

Guidance for Semantic Video Search on AWS

Steps 8-11



8 A foundation model (FM) in **Amazon Bedrock** generates image descriptions from the frames data and detected text or celebrities in the frames. The FM creates a description of the video shot based on a sequence of descriptions of its associated frames.

9 An embedding model in **Amazon Bedrock** generates the embeddings of a video's descriptions. **Amazon OpenSearch Service** stores the embeddings and other related information in a vector database.

10 An embedding model in **Amazon Bedrock** generates the embedding of the users' query, which is then used to perform semantic search for the videos from the **OpenSearch Service** vector database.

11 **Amazon DynamoDB** tables store profiling and video indexing job metadata to keep track of the jobs' status and other relevant information.

AWS Reference Architecture