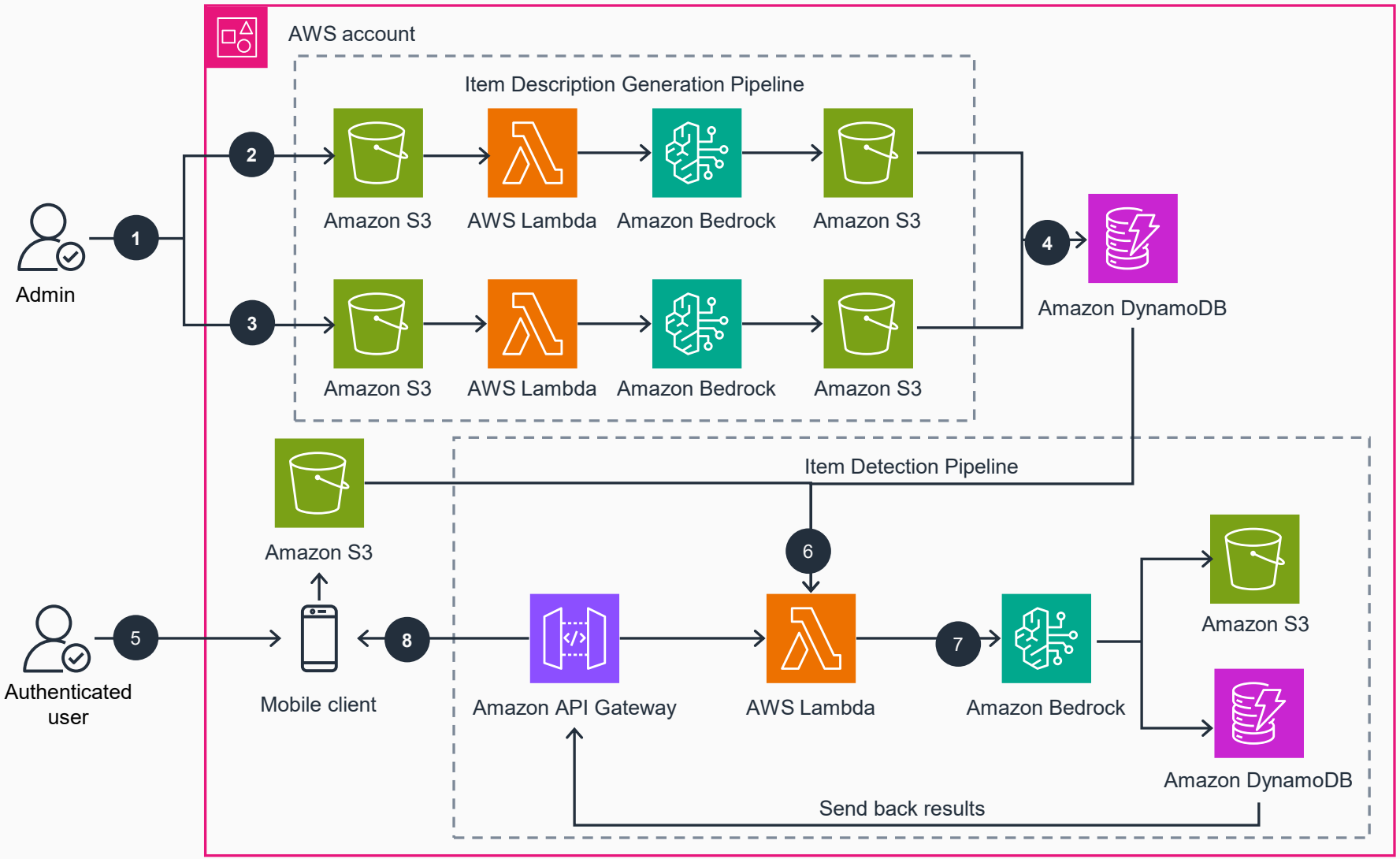
Guidance for Generative AI-Powered Visual Inspection on AWS

This architecture diagram illustrates how to effectively support on-site visual inspection and verification on AWS. It shows the key components and their interactions, providing an overview of the architecture's structure and functionality.



1. Admin uploads images of products and bill of material (BOM) to **Amazon Simple Storage Service (Amazon S3)** bucket.
2. Item description generation pipeline **Lambda** is triggered to process each reference image with **Amazon Bedrock** (a fully managed service providing access to foundation models, enables the solution to leverage advanced AI capabilities) Claude Sonnet model and generates a reference rich and detailed description of the image.
3. Descriptions are combined with BOM data to generate module specific detection rules. Additional rules for common false positives patterns are generated for improved accuracy.
4. Descriptions and specific detection rules are stored in **Amazon S3** and synced to **Amazon DynamoDB** table.
5. The detection pipeline is triggered when a user upload the image to **Amazon S3** while performing an on-site visual inspection.
6. A **Lambda** function is triggered and prepares the context with the uploaded image and target module and description of items belonging to the module retrieved from **DynamoDB,** retrieving relevant reference descriptions and detection rules from **DynamoDB** to provide context to the model (RAG approach).
7. **Amazon Bedrock Nova Pro** perform the detection task and output results to S3 and **DynamoDB**.
8. Detection results are retrieved by the mobile client through **Amazon API Gateway**, with all the items detected. User can move to next module for verification or restart the detection by taking a different picture to detect the missing items.

AWS Reference Architecture