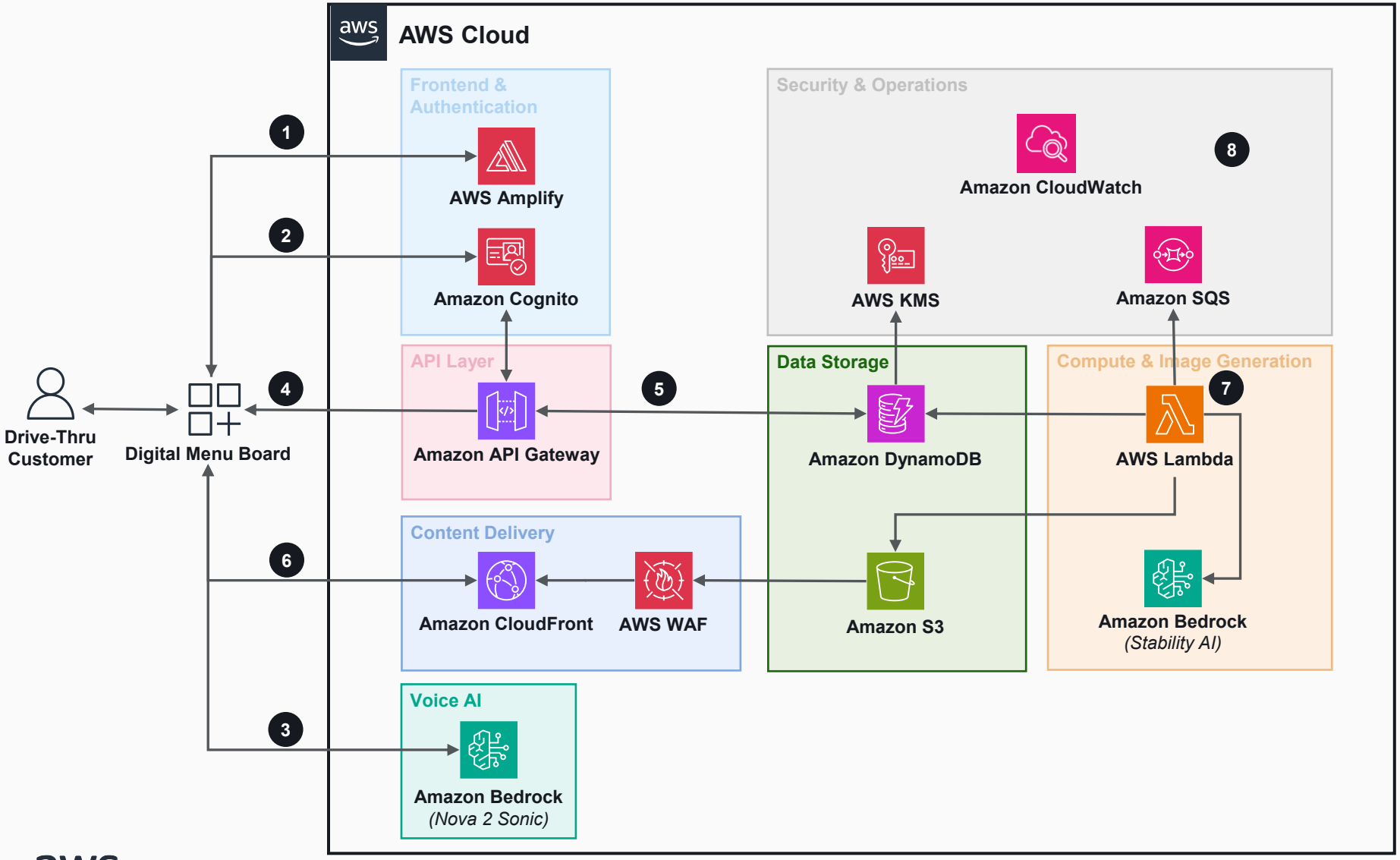
Guidance for Drive Thru Voice AI using Amazon Bedrock

This architecture diagram shows how Amazon Nova 2 Sonic enables near real-time voice AI ordering for quick-service restaurant drive-thru using AWS serverless services.



1. You access the digital menu board hosted on **AWS Amplify**, which loads the React-based ordering interface.
2. **Amazon Cognito** manages temporary AWS credentials with scoped IAM permissions and secures access to backend services through token-based authorization.
3. The Digital Menu Board uses the AWS SDK BidirectionalStreaming API to establish a SigV4-signed WebSocket session with **Amazon Bedrock**, a fully managed service with built-in security, privacy, and responsible AI. The app passes tool definitions and a system prompt to **Amazon Nova 2 Sonic**, enabling the model to orchestrate tool calls and process streaming audio.
4. **Amazon Nova 2 Sonic** processes your streaming audio and invokes tools defined at the application level. The app's tool router captures model requests with parameters to manipulate the UI and trigger API calls.
5. You initiate voice input via a button or sensor trigger. **Amazon API Gateway** routes tool requests to **Amazon DynamoDB** tables for menu, cart, order, loyalty, and chat data.
6. **Amazon CloudFront** delivers AI-generated menu images stored in **Amazon Simple Storage Service (Amazon S3)**, protected by **AWS WAF**, a web application firewall that filters malicious traffic and blocks unauthorized access.
7. **AWS Lambda** invokes Stability AI's Stable Image Core model through **Amazon Bedrock** to generate menu images, stores the image binary in **Amazon S3**, and updates the image reference in **Amazon DynamoDB**. (This happens once during the solution deployment)
8. **AWS KMS** encrypts DynamoDB tables. **Amazon Simple Queue Service (Amazon SQS)** dead letter queues capture failed **AWS Lambda** invocations for retry and alerting. **Amazon CloudWatch** provides centralized monitoring and logging.

Reviewed for technical accuracy April 2026

AWS Reference Architecture