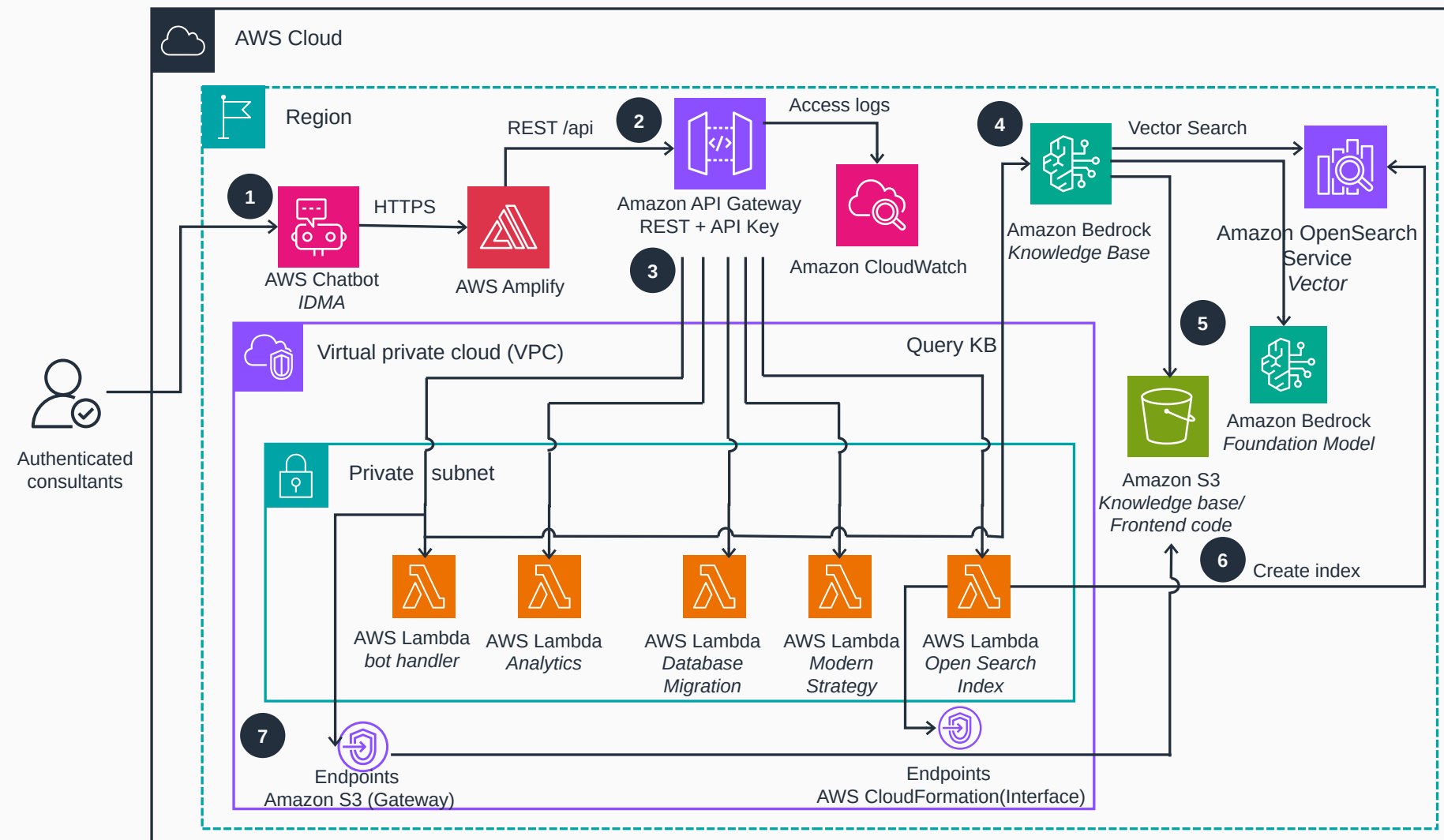


Guidance for Intelligent Data Modernization Assistant on AWS

This architecture diagram shows how to build an AI-powered assistant that uses RAG-based generative AI to streamline end-to-end data platform migrations, analytics workload modernization, and automated strategy generation for databases, data lakes, and catalogs on AWS.



- 1 Solution architects and consultants access the **Intelligent Data Migration Assistant (IDMA) Chatbot** application on **AWS Amplify**, which serves the React-based frontend UI.
- 2 The IDMA Chatbot sends requests through **HTTPS** to **Amazon API Gateway** (REST API secured with API Key) for backend processing.
- 3 API Gateway routes requests to **AWS Lambda functions** (bot-handler, database-migration, analytics, modern-strategy) running inside **Amazon Virtual Private Cloud (VPC)** private subnets based on the API endpoint path.
- 4 Lambda functions invoke **Amazon Bedrock Knowledge Base** using the RAG (Retrieve and Generate) pattern to query curated migration documents and generate AI-powered responses.
- 5 Bedrock Knowledge Base uses **Amazon Titan Embed v2** (Foundation Model) to convert queries into vector embeddings and performs similarity search against **Amazon OpenSearch Serverless** to retrieve relevant document context.
- 6 A dedicated Lambda (CloudFormation custom resource) creates the **KNN vector index** schema in OpenSearch Serverless, which Amazon Bedrock Knowledge Base uses to store and query vector embeddings.
- 7 Lambda functions access **Amazon Simple Storage Service (Amazon S3)** bucket via the **S3 Gateway Endpoint** within the VPC for secure, private connectivity.

