

RAG e Base de Conhecimento

O que é uma Base de Conhecimento

No Contexto AWS:

• **Definição:** Um repositório organizado de informações que pode ser usado por modelos de IA para responder perguntas com dados confidenciais.

• **Serviços AWS relevantes:**

• **Amazon Bedrock Knowledge Bases:** permite integrar dados privados em um fluxo RAG.

• **Amazon Kendra:** motor de busca inteligente que indexa e recupera informações de fontes internas.

• **Objetivo para prova:** Dever identificar quando usar uma Base de Conhecimento para melhorar respostas da IA.

O que é RAG (Geração Aumentada por Recuperação)

- **Definição:** Técnica onde a IA primeiro **recupera** informações relevantes de uma Base de Conhecimento e depois **gera** a resposta usando um modelo de linguagem.

• Fluxo típico:

- 1- O usuário faz uma pergunta.
- 2- O sistema busca (retrieval) dados em fontes confiáveis.
- 3- O modelo de IA usa esses dados para gerar (generation) a resposta.

- **Na AWS:** implementado via **Amazon Bedrock + Knowledge Bases + Conectores** (ex: Kendra, S3, Bancos de dados).

- **Benefício para pessoa:** Reduz alucinações e aumenta precisão.

Como as coisas se conectam.

- O RAG só funciona bem se houver uma Base de conhecimento organizada e acessível.

- A base fornece dados, e o RAG é a sobre-estrutura que combina **Busca + geração**.
- No AWS Bedrock, a **Knowledge Base** é configurada para indexar dados e servir como fonte para o modelo.

Casos de uso práticos (AIK - CO1)

- Atendimento ao cliente com FAQ dinâmico (Usando dados internos sempre atualizados).
- Pesquisa interna em documentos corporativos.
- Assistentes virtuais que respondem com Base em políticas e manuais internos.
- Análise de contratos ou relatórios técnicos com suporte a perguntas em linguagem natural.

