Advanced Generative AI Development on AWS

- **Nível:** Avançado
- **Duração:** 21h

Sobre o curso

O curso Advanced Generative AI Development on AWS foi desenhado para programadores que procuram dominar a implementação de soluções de generative AI prontas para produção na AWS.

O curso responde às necessidades de organizações que iniciam o seu percurso em generative AI e aborda a forma de construir estratégias abrangentes de generative AI alinhadas com objetivos empresariais mais amplos.

Esta formação avançada de 3 dias, ministrada por um formador, **desenvolve competências em toda a stack de generative AI – desde foundation models até padrões de integração empresarial**.

Adicionalmente, irás aprender sobre técnicas avançadas de processamento de dados, implementação de vector databases e retrieval augmentation, prompt engineering avançado e governance, sistemas de agentic AI e integração de ferramentas, medidas de AI safety e security, otimização de desempenho e estratégias de gestão de custos, soluções abrangentes de monitorização e observabilidade, e frameworks de testes e validação.

A estrutura do curso segue o modelo comprovado da AWS para adoção de generative AI, evoluindo da experimentação para implementações prontas para produção.

Destinatários

- Programadores de software
- Profissionais técnicos

Objetivos

- Desenvolver soluções de generative AI prontas para produção utilizando serviços AWS que cumpram requisitos empresariais de segurança, escalabilidade e fiabilidade
- Avaliar e selecionar foundation models adequados para casos de utilização empresariais específicos, incluindo benchmarking de desempenho e implementação de arquiteturas de seleção dinâmica de modelos
- Desenhar e implementar sistemas de foundation models com circuit breakers, implementação cross-

region e estratégias de degradação
- Criar pipelines abrangentes de processamento de dados para inputs multimodais, incluindo workflows de validação e técnicas de otimização
- Implementar soluções sofisticadas de vector databases utilizando Amazon Bedrock Knowledge Bases, OpenSearch e abordagens híbridas para retrieval augmentation eficaz
- Criar e gerir frameworks avançados de prompt engineering, incluindo chain-of-thought reasoning e sistemas de prompt governance à escala empresarial
- Explicar os componentes de frameworks de Agentic AI e do Amazon Bedrock AgentCore
- Implementar controlos abrangentes de AI safety e security, incluindo filtragem de conteúdos, preservação da privacidade e mecanismos de testes adversariais
- Otimizar o desempenho e gerir custos através de estratégias de eficiência de tokens, implementações de batching e sistemas de caching inteligente
- Desenhar e implementar soluções abrangentes de monitorização e observabilidade para aplicações baseadas em foundation models
- Criar frameworks sistemáticos de testes e validação para garantia contínua da qualidade de aplicações de AI
- Integrar soluções de generative AI em ambientes empresariais utilizando padrões arquiteturais seguros, conformes e escaláveis

---

Pré-requisitos

Recomenda-se que os participantes deste curso tenham:

- Realizado ou conhecimentos equivalentes ao curso AWS Technical Essentials
- Realizado ou conhecimentos equivalentes ao curso Generative AI Essentials on AWS
- 2 ou mais anos de experiência no desenvolvimento de aplicações de nível de produção na AWS ou com tecnologias opensource,
  experiência geral em AI/ML ou data engineering
- 1 ano de experiência prática na implementação de soluções de generative AI

Programa

- Seleção e Configuração de Foundation Models
- Processamento Avançado de Dados para Foundation Models
- Vector Databases e Retrieval Augmentation
- Prompt Engineering e Governance
- Implementação de Frameworks de Agentic AI com Amazon Bedrock AgentCore
- AI Safety e Security
- Otimização de Desempenho e Gestão de Custos
- Monitorização e Observabilidade para Generative AI
- Testes, Validação e Melhoria Contínua
- Padrões de Integração Empresarial
- Encerramento do Curso

Seleção e Configuração de Foundation Models

- Framework empresarial de avaliação de foundation models
- Padrões de arquitetura para seleção dinâmica de modelos
- Desenhos resilientes de sistemas de foundation models
- Otimização de custos e modelação económica

Processamento Avançado de Dados para Foundation Models

- Validação abrangente de dados e garantia de qualidade
- Pipelines de processamento de dados multimodais
- Otimização de inputs e melhoria de desempenho

Vector Databases e Retrieval Augmentation

- Arquitetura empresarial de vector databases
- Processamento avançado de documentos e estratégias de chunking
- Implementação de sistemas sofisticados de retrieval
- Hands-on Lab: Desenvolver aplicações de Retrieval Augmented Generation (RAG) com Amazon Bedrock Knowledge Bases

Dia 2

Prompt Engineering e Governance

- Frameworks avançados de prompt engineering
- Sistemas complexos de orquestração de prompts
- Prompt governance e gestão à escala empresarial

Hands-on Lab: Desenvolver padrões de conversação com Amazon Bedrock APIs

Implementação de Frameworks de Agentic AI com Amazon Bedrock AgentCore

- Frameworks de Agentic AI
- Amazon Bedrock AgentCore

AI Safety e Security

- Implementação abrangente de content safety
- Arquitetura de AI com preservação da privacidade
- Frameworks de AI governance e conformidade

Dia 3

Otimização de Desempenho e Gestão de Custos

- Eficiência de tokens e otimização de custos
- Arquitetura de sistemas de alto desempenho
- Implementação de sistemas de caching inteligente
- **Hands-on Lab:** Criar Gen AI segura e responsável com Guardrails for Amazon Bedrock

Monitorização e Observabilidade para Generative AI

- Sistemas de monitorização de foundation models
- Gestão de impacto e valor para o negócio
- Troubleshooting e diagnóstico específicos de AI

Testes, Validação e Melhoria Contínua

- Frameworks abrangentes de avaliação de AI
- Garantia de qualidade e melhoria contínua
- Avaliação e otimização de sistemas RAG

Padrões de Integração Empresarial

- Conectividade empresarial e arquitetura de integração
- Acesso seguro e gestão de identidades
- Implementações cross-environment e híbridas

Encerramento do Curso

- Próximos passos e recursos adicionais
- Resumo do curso