AKADEMIA
FORMATION

PROGRAMME DE FORMATION

Claude sur AWS Bedrock — Intégration Cloud Amazon

DURÉE	FORMAT	PUBLIC	CERTIFICATION
2 jours	**Inter · Intra**	**Tout collaborateur**	**Incluse**
14 heures	Présentiel ou distanciel		

À PROPOS DE LA FORMATION

Claude sur AWS Bedrock — Intégration Cloud Amazon.

DURÉE	FORMAT	PÉDAGOGIE	CERTIFICATION
2 jours	**Inter · Intra**	**Active**	**Incluse**
14 heures	Présentiel ou distanciel	Petits groupes	Attestation délivrée

Objectifs pédagogiques

- Configurer l'accès aux modèles Claude sur Amazon Bedrock (formulaire FTU, IAM, régions)
- Maîtriser les trois méthodes d'appel : SDK Anthropic pour Bedrock, Boto3 InvokeModel et Converse API
- Implémenter le streaming, le tool use et l'extended thinking via Bedrock
- Déployer des Guardrails Bedrock pour le filtrage de contenu, la détection PII et la protection anti-jailbreak
- Construire un pipeline RAG avec Bedrock Knowledge Bases (S3, embedding, recherche vectorielle)
- Créer des agents autonomes avec Bedrock Agents intégrés à Lambda et Step Functions
- Optimiser les coûts : endpoints globaux vs régionaux, prompt caching, Batch API et choix de modèle

PROGRAMME

Huit modules progressifs pour monter en compétences.

MODULE

01.

1H30H

Pourquoi Bedrock : positionnement et avantages

CONTENU PÉDAGOGIQUE

- Trois voies d'accès à Claude : API directe Anthropic, Amazon Bedrock, Google Vertex AI
- Avantages Bedrock : facturation AWS consolidée, conformité SOC/HIPAA/FedRAMP, IAM natif
- Arbre de décision : quand choisir Bedrock vs API directe selon les contraintes (conformité, latence, coûts)
- Modèles disponibles sur Bedrock : Opus 4.6, Sonnet 4.6, Haiku 4.5 avec Model IDs et disponibilité régionale
- Fenêtres de contexte : 1M tokens (Opus/Sonnet 4.6) vs 200K tokens (autres modèles), limite payload 20 Mo
- Exercice : Comparer les 3 plateformes sur 3 cas d'usage entreprise et choisir la plus adaptée

MODULE

02.

2H

Setup et première requête Bedrock

CONTENU PÉDAGOGIQUE

- Prérequis : compte AWS actif, AWS CLI 2.13.23+, activation Bedrock, formulaire First Time Use (FTU)
- Permissions IAM : politique minimale pour invoquer les modèles, vérification avec aws sts get-caller-identity
- Installation des SDKs : pip install anthropic[bedrock] (Python), npm install @anthropic-ai/bedrock-sdk (TS)
- Requête avec le SDK Anthropic : AnthropicBedrock, credentials (access key, secret, session token, region)
- Requête avec Boto3 : bedrock-runtime, InvokeModel, anthropic_version bedrock-2023-05-31
- Endpoints globaux (préfixe global., routage dynamique, pas de surcharge) vs régionaux (CRIS, +10%)
- Bearer Token authentication : simplifier l'accès dans les environnements d'entreprise
- Atelier pratique : Configurer l'environnement AWS, lister les modèles et exécuter des requêtes via les 3 méthodes

MODULE

03.

1H45H

API Bedrock avancée : streaming, tool use et extended thinking

CONTENU PÉDAGOGIQUE

- Streaming sur Bedrock : format AWS event-stream, implémentation Python et TypeScript
- Tool Use via la Converse API (recommandée) : toolSpec, inputSchema, gestion tool_use et tool_result
- Tool Use via InvokeModel : format natif Anthropic sur Bedrock, comparaison avec Converse API
- Extended Thinking sur Bedrock : activation, budget_tokens, mode adaptive (Opus 4.6/Sonnet 4.6)
- Vision et PDF : analyse d'images, traitement de documents, citations via Converse API
- Prompt caching sur Bedrock : cache_control, TTL 5 min/1h, tarification identique à l'API directe
- Exercice : Développer un assistant avec tool use et streaming qui interroge une API métier via Bedrock

MODULE

04.

1H45H

Bedrock Guardrails : sécurité et conformité

CONTENU PÉDAGOGIQUE

- Les 6 politiques de sauvegarde : filtres de contenu, détection d'attaques prompt, sujets interdits, redaction PII, grounding contextuel, automated reasoning
- Création et publication d'un Guardrail dans la console Amazon Bedrock
- Intégration API : headers X-Amzn-Bedrock-GuardrailIdentifier et GuardrailVersion
- API ApplyGuardrail : filtrage de contenu indépendamment du modèle, sans invoquer Claude
- Cross-Region inference : compatibilité guardrails avec les profils d'inférence cross-region
- Cas d'usage : conformité RGPD (masquage PII), chatbot client avec sujets interdits, protection anti-jailbreak
- Atelier pratique : Créer un Guardrail complet (PII + sujets interdits + anti-jailbreak) et le tester en production

MODULE

05.

1H45H

Bedrock Knowledge Bases : RAG entièrement géré

CONTENU PÉDAGOGIQUE

- Architecture du flux RAG Bedrock : ingestion → embedding → stockage vectoriel → retrieval → augmentation → génération
- Sources de données : S3, Confluence, Salesforce, SharePoint, Web Crawler
- Modèles d'embedding : Amazon Titan Embeddings pour la vectorisation des chunks
- Stockages vectoriels : OpenSearch Serverless, Aurora, Neptune, MongoDB, Pinecone, Redis Enterprise Cloud
- APIs clés : Retrieve API (résultats bruts avec visuels) et RetrieveAndGenerate API (réponse directe)
- Implémentation Python : bedrock-agent-runtime, knowledgeBaseId, modelArn avec Claude
- Exercice : Créer une Knowledge Base depuis un bucket S3, indexer des documents et interroger avec Claude

MODULE

06.

1H45H

Bedrock Agents et intégration Lambda/Step Functions

CONTENU PÉDAGOGIQUE

- Bedrock Agents : agents autonomes avec raisonnement, planification et exécution d'actions
- Tool Use (Function Calling) sur Bedrock : définition d'outils avec la Converse API
- Intégration Lambda : connecter les agents à des fonctions serverless pour l'exécution d'actions métier
- AgentCore : exécution autonome jusqu'à 8 heures avec scaling automatique
- Server-side vs client-side tool calling : quand déléguer l'exécution à Bedrock vs la gérer côté client
- Orchestration Step Functions : combiner Claude avec d'autres services AWS dans des workflows complexes
- Atelier pratique : Construire un agent Bedrock avec 3 outils Lambda (recherche KB, appel API, écriture S3)

MODULE

07.

1H45H

Optimisation des coûts et performance sur Bedrock

CONTENU PÉDAGOGIQUE

- Tarification Bedrock : identique à l'API directe sur les endpoints globaux, +10% sur régionaux
- Batch API sur Bedrock : -50% sur les prix, jusqu'à 100 000 requêtes par batch
- Prompt caching : écriture 5 min (1.25x) et 1h (2x), lecture à 0.1x = économie de 90%
- Stratégie de choix de modèle : Haiku 4.5 pour le volume, Sonnet 4.6 pour l'équilibre, Opus 4.6 pour la qualité
- Provisioned throughput vs on-demand : quand réserver de la capacité dédiée
- Fonctionnalités exclusives API directe non disponibles sur Bedrock : web search, web fetch, code execution
- Exercice : Calculer le TCO d'un workflow et optimiser avec caching + batch + modèle adapté

MODULE

08.

1H45H

Architecture de production et bonnes pratiques

CONTENU PÉDAGOGIQUE

- Architecture de référence : API Gateway → Lambda → Bedrock, DynamoDB pour le state, S3 pour les documents
- Monitoring CloudWatch : métriques d'invocation, latence, tokens, taux d'erreurs, alertes de coûts
- Journalisation des invocations Bedrock : prompts et complétions, rétention 30 jours recommandée
- Sécurité : IAM least privilege, VPC endpoints pour Bedrock, chiffrement at-rest et in-transit
- Haute disponibilité : endpoints globaux pour la résilience, fallback entre régions
- SDKs disponibles : Python, TypeScript, Go, Java, C#, PHP, Ruby — exemples comparés
- Atelier final : Concevoir et documenter une architecture Bedrock complète avec Guardrails, Knowledge Bases et monitoring

PASSONS À L'ACTION

Construisons ensemble votre session sur-mesure.

Dites-nous vos contraintes (format, lieu, dates, nombre de participants) et recevez une proposition personnalisée sous 24 heures ouvrées.

Akademia Formation

SERVICE ADMINISTRATION DES VENTES

adv@akademiaformation.com

www.akademiaformation.com

Devis personnalisé

RÉPONSE SOUS 24 H OUVRÉES

Format inter · intra

Présentiel ou distanciel

FIN DU PROGRAMME