

Certifiante

Formation Développer des solutions Cloud IA sur Azure (AI-200)

Réf. SF34743

Suivez le cours officiel Microsoft AI-200T00 pour construire les services back-end, les bases vectorielles et l'observabilité dont vos applications d'IA ont besoin pour tenir en production sur Azure.

5 j
Durée

3 750 € HT
Inter / participant

0
Sessions à venir

Objectifs pédagogiques

- ✓ Déployer et exploiter des applications conteneurisées sur Azure Container Registry et Azure App Service, sidecars compris
- ✓ Déployer, mettre à l'échelle et dépanner des applications sur Azure Container Apps et Azure Kubernetes Service
- ✓ Concevoir et interroger des bases de données documentaires et relationnelles pour charges IA avec Azure Cosmos DB for NoSQL et Azure Database for PostgreSQL
- ✓ Implémenter la recherche vectorielle et des patterns RAG avec pgvector, Cosmos DB et Azure Managed Redis
- ✓ Optimiser les performances et le coût des requêtes et des recherches vectorielles par l'indexation, les niveaux de cohérence, le cache et le pooling de connexions
- ✓ Développer des back-ends IA événementiels et serverless avec Azure Service Bus, Event Grid, Azure Functions et Durable Functions
- ✓ Sécuriser les secrets et centraliser la configuration avec Azure Key Vault et App Configuration, et instrumenter les applications avec OpenTelemetry et Azure Monitor
- ✓ Préparer l'examen de certification AI-200 — Microsoft Certified: Azure AI Cloud Developer Associate

Public visé & prérequis

Public visé Développeurs

Prérequis- Pratique de la programmation **Python** — les travaux pratiques du cours sont en Python - Expérience du développement d'applications sur Azure, correspondant au périmètre historique du cours AZ-204 : App Service, Azure Functions, stockage, authentification, API REST et architectures événementielles - Pratique de la conteneurisation (images, registres, Docker) et fondamentaux de Kubernetes : pods, services, manifestes - Notions d'embeddings et de recherche par similarité, et compréhension du principe du RAG — le cours apprend à stocker et interroger des vecteurs, non à les produire - Bases du SQL et des structures de documents JSON - Aisance avec la ligne de commande et l'Azure CLI - Lecture de l'anglais technique (supports de cours officiels)

Programme détaillé

Hébergement d'applications conteneurisées sur Azure

- Stocker et gérer des conteneurs dans Azure Container Registry : hiérarchie du registre, build d'images avec ACR Tasks, stratégies de tags et de versions
- Déployer des conteneurs sur Azure App Service : sources de conteneurs, paramètres d'exécution, configuration applicative et diagnostics
- Exécuter des applications IA avec sidecars sur Azure App Service : composants IA couplés au conteneur principal, pulls d'images sécurisés, communication locale et fichiers partagés

Activités pratiques :

- Build d'une image dans Azure Container Registry avec ACR Tasks et application d'une stratégie de tags
- Déploiement d'un conteneur applicatif sur Azure App Service avec variables d'environnement et secrets
- Ajout d'un sidecar à une application Linux et vérification de la communication entre conteneurs

Déploiement et mise à l'échelle sur Azure Container Apps

- Déployer des conteneurs sur Azure Container Apps : environnements, configuration d'exécution, authentification au registre et vérification du déploiement
- Gérer les conteneurs dans Azure Container Apps : mise à jour d'images, gestion des révisions, ajustement des ressources, sondes de santé et logs
- Mettre à l'échelle les conteneurs dans Azure Container Apps : règles HTTP, TCP, CPU et mémoire, mise à l'échelle événementielle avec KEDA, modes de révision et répartition du trafic

Activités pratiques :

- Déploiement d'une application dans un environnement Container Apps avec authentification au registre
- Publication d'une nouvelle révision et bascule progressive du trafic
- Configuration d'une règle de mise à l'échelle événementielle avec KEDA

Déploiement et supervision sur Azure Kubernetes Service

- Déployer des applications sur Azure Kubernetes Service : manifestes de déploiement et exposition des applications via les Services Kubernetes
- Configurer des applications sur Azure Kubernetes Service : externalisation de la configuration par ConfigMaps, sécurisation des paramètres sensibles par Secrets, stockage persistant
- Surveiller et dépanner des applications sur Azure Kubernetes Service : santé applicative, logs et métriques, diagnostic des pods et des Services, vérification de la connectivité

Activités pratiques :

- Rédaction d'un manifeste de déploiement et exposition de l'application par un Service
- Externalisation de la configuration par ConfigMap et montage d'un volume persistant
- Diagnostic d'un pod en échec à partir des logs et des événements du cluster

Solutions IA avec Azure Cosmos DB for NoSQL

- Créer des requêtes pour Azure Cosmos DB for NoSQL : connexion par SDK, opérations sur les items et requêtes SQL
- Mettre en œuvre la recherche vectorielle sur Azure Cosmos DB for NoSQL : stockage des embeddings, fonction VectorDistance, filtres de métadonnées, recherche hybride et synchronisation par change feed
- Optimiser les performances des requêtes pour Azure Cosmos DB for NoSQL : index de plage et composites, types d'index vectoriels, niveaux de cohérence et consommation de RU

Activités pratiques :

- Connexion par SDK et interrogation d'un conteneur de documents en SQL
- Recherche de similarité avec VectorDistance combinée à un filtre de métadonnées
- Mise en place d'un change feed processor pour maintenir des embeddings à jour

Solutions IA avec Azure Database for PostgreSQL

- Créer et interroger des bases de données avec Azure Database for PostgreSQL : conception de schémas, requêtes efficaces, intégration applicative avec authentification Microsoft Entra
- Mettre en œuvre la recherche vectorielle avec Azure Database for PostgreSQL : extension pgvector, stockage des embeddings, index vectoriels et patterns de récupération sémantique pour le RAG
- Optimiser la recherche vectorielle dans Azure Database for PostgreSQL : paramètres de configuration, choix des index, organisation des données, montée en charge et pooling de connexions

Activités pratiques :

- Conception d'un schéma et connexion applicative sécurisée par identité managée
- Stockage d'embeddings avec pgvector et recherche de similarité selon plusieurs métriques de distance
- Comparaison de deux stratégies d'index vectoriel sur un même jeu de requêtes

Azure Managed Redis pour les solutions IA

- Mettre en œuvre des opérations sur les données dans Azure Managed Redis : fonctionnalités du service, bonnes pratiques des bibliothèques clientes, cache, expiration et invalidation
- Mettre en œuvre la messagerie d'événements avec Azure Managed Redis : pub/sub pour la diffusion de notifications et Redis Streams pour le traitement asynchrone fiable
- Mettre en œuvre le stockage vectoriel dans Azure Managed Redis : index vectoriels, interrogation des embeddings, types de vecteurs et stratégies d'indexation

Activités pratiques :

- Mise en cache de réponses d'inférence avec politique d'expiration et invalidation
- Coordination d'un pipeline de traitement multi-étapes avec Redis Streams
- Création d'un index vectoriel et exécution d'une recherche par similarité

Intégration des services back-end pour les solutions IA

- Mettre en file d'attente et traiter les opérations d'IA avec Azure Service Bus : queues, topics et souscriptions, structuration des messages, consommateurs concurrents et dead-letter queues
- Développer des workflows d'IA pilotés par les événements avec Azure Event Grid : souscriptions d'événements, CloudEvents, filtres, politiques de livraison et nouvelles tentatives, événements personnalisés
- Créer des back-ends d'IA sans serveur avec Azure Functions : déclencheurs et liaisons, endpoints d'inférence, processeurs d'événements, configuration et déploiement des fonction apps
- Orchestrer des workflows d'IA durables avec Azure Durable Functions : orchestrations déterministes, traitement parallèle, attente de décisions humaines et reprise sur erreur

Activités pratiques :

- Mise en file de requêtes d'inférence avec Service Bus et traitement des échecs par dead-letter queue
- Routage d'événements vers plusieurs handlers avec Event Grid et publication d'un événement personnalisé
- Développement d'une API d'inférence serverless avec Azure Functions
- Orchestration d'un traitement long avec approbation humaine via Durable Functions

Secrets, configuration et observabilité des applications IA

- Gérer les secrets d'application avec Azure Key Vault : organisation du coffre, récupération par SDK avec identité managée, versions et rotation, stratégies de cache
- Gérer les paramètres d'application avec Azure App Configuration : centralisation des paramètres, labels par environnement, feature flags et références aux secrets Key Vault
- Instrumenter une application avec OpenTelemetry : traces distribuées, spans personnalisés, export de la télémétrie vers Application Insights
- Analyser la télémétrie des applications avec les journaux et les métriques : requêtes KQL, patterns d'erreurs et tendances de performance, tableaux de bord, workbooks et alertes

Activités pratiques :

- Récupération d'un secret Key Vault par identité managée et mise en place d'une rotation
- Centralisation de la configuration avec labels et feature flags, et référencement d'un secret Key Vault
- Instrumentation d'une application distribuée avec OpenTelemetry et lecture d'une trace de bout en bout
- Rédaction de requêtes KQL pour identifier un goulot d'étranglement et configuration d'une alerte

Préparation à la certification

- Modalités de l'examen : examen AI-200 surveillé, 120 minutes, seuil de réussite 700/1000, composants interactifs possibles, disponible en français
- Examen blanc : passage d'un test en conditions réelles
- Stratégies de réponse : gestion du temps, lecture des questions, élimination des distracteurs
- Révision ciblée : synthèse des points clés par domaine de la trame — solutions conteneurisées, services de données pour l'IA, connexion aux services Azure, sécurité et supervision
- Planification : inscription à l'examen auprès de Pearson VUE et conseils de préparation post-formation

Activités pratiques :

- Examen blanc chronométré avec correction commentée
- Atelier de révision par domaine avec quiz

Pourquoi choisir Sparks Formation ?

20 ans d'expertise IT, des formateurs experts et des sessions garanties contre l'annulation.

20 ans
d'expertise

1 068+
formations

11 178
apprenants formés/an

+67
Net Promoter Score

Avis des participants

4,9/5 — 110 avis

Selim B. — Très bonne formation, très instructif même après quelques années de Python dans les bras. Une très petite équipe de stagiaires qui facilite les échanges. Je redoutais le mode...

Alexis C. — Anthony explique bien les concepts attendus et prend le temps de répondre aux questions et aux différents problèmes rencontrés. Le support présenté créé par lui-même est différent des...

Ils nous font confiance

Cette année notre filière Langages & Écosystèmes Backend a accompagné plus de 292 entreprises dont :

SOPRA STERIA RENAULT SAS ORANGE CAMPUS TECH ORANGE INNOVATION SCHOOL STMICROELECTRONICS GRANT THORNTON FRANCE
ORANGE BUSINESS RTE SII ILE DE FRANCE THALES GLOBAL SERVICES +282 autres

Financement & certifications

Cette formation est éligible aux financements OPCO et FNE-Formation. Nos équipes vous accompagnent dans le montage de votre dossier.



Réservez votre place dès maintenant

20 ans d'expertise · 1 000+ formations · NPS +67 · réponse sous 24h ouvrées

0 805 950 800

demande@sparks-formation.com