Développer des solutions cloud IA sur Azure

Formation officielle Microsoft Azure AI-200 : Developing AI Solutions for Microsoft Azure

Classe à distance
Développement

Code :
AI200

Durée :
5 jour(s) (35,00 heures)

Exposés : **40 %**
Cas pratiques : **40 %**
Echanges d'expérience : **20 %**

Inter-entreprises :
Prochaines sessions disponibles sur notre site web.
Tarif : 3 130,00 € HT / participant

Intra-entreprise :
Tarifs et dates sur demande.

DESCRIPTION

Cette formation remplace la formation « AZ204 : Azure pour les développeurs ».

Cette formation vous permettra d'apprendre à créer, superviser et dépanner des solutions d'intelligence artificielle sur Microsoft Azure.

Vous découvrirez comment déployer des applications à l'aide des modèles de calcul et de conteneurisation Azure, créer des API serverless avec Azure Functions et intégrer différents services via des architectures orientées événements telles qu'Azure Service Bus ou Event Grid.

Vous apprendrez également à exploiter les services de données Azure adaptés aux charges de travail IA, notamment en concevant et interrogeant des solutions avec Cosmos DB pour NoSQL, Azure Database for PostgreSQL avec pgvector, ainsi qu'Azure Managed Redis pour la mise en cache, le streaming et la recherche vectorielle.

Enfin, vous connecterez des services, orchestrerez des workflows d'IA et concevrez des applications IA sécurisées, évolutives et observables sur Azure.

OBJECTIFS PEDAGOGIQUES

Les participants seront en mesure de :

- Implémenter l'hébergement d'applications conteneur sur Azure
- Déployer et gérer des applications sur Azure Container Apps
- Déployer et surveiller des applications sur Azure Kubernetes Service
- Développer des solutions IA avec Azure Cosmos DB pour NoSQL
- Développer des solutions IA avec Azure Database pour PostgreSQL
- Améliorer les solutions IA avec Azure Managed Redis
- Intégrer des services back-end pour les solutions IA
- Gérer les secrets et la configuration des applications pour les solutions IA
- Observer et diagnostiquer des applications sur Azure

PUBLIC CIBLE

Développeurs

PRE-REQUIS

Pour suivre cette formation, vous devez avoir suivi la formation « AZ-900 : Les fondamentaux Azure » et « AI-901 : Présentation de l'IA dans Azure » ou avoir une expérience opérationnelle équivalente.

Les supports de cours et les travaux pratiques sont en anglais. Un niveau d'anglais B1 est recommandé pour suivre cette formation.

Retrouvez les niveaux de langue sur ce lien : Classification des niveaux de langue

Il est fortement recommandé de suivre ce cours sur un ordinateur et de disposer d'un double écran pour plus de confort.

METHODE PEDAGOGIQUE

Formation avec apports théoriques, échanges sur les contextes des participants et retours d'expérience pratique des formateurs, complétés de travaux pratiques et de mises en situation.

PROFIL DES INTERVENANTS

Cette formation est dispensée par un·e ou plusieurs consultant·es d'OCTO Technology ou de son réseau de partenaires, expert·es reconnus des sujets traités.

Le processus de sélection de nos formateurs et formatrices est exigeant et repose sur une évaluation rigoureuse leurs capacités techniques, de leur expérience professionnelle et de leurs compétences pédagogiques.

MODALITÉS D'ÉVALUATION ET FORMALISATION À L'ISSUE DE LA FORMATION

L'évaluation des acquis se fait tout au long de la session au travers des ateliers et des mises en pratique. Afin de valider les compétences acquises lors de la formation, un formulaire d'auto-positionnement est envoyé en amont et en aval de celle-ci. Une évaluation à chaud est également effectuée en fin de session pour mesurer la satisfaction des stagiaires et un certificat de réalisation leur est adressé individuellement.

PROGRAMME PEDAGOGIQUE DETAILLE

STOCKER ET GÉRER DES CONTENEURS DANS AZURE CONTAINER REGISTRY

- Expliquer comment Azure Container Registry organise des images à l'aide de registres, de référentiels et d'artefacts
- Créer et gérer des images conteneur dans le cloud à l'aide d'ACR Tasks

- Implémenter des stratégies de balisage et de contrôle de version pour les déploiements de conteneurs fiables
- Utiliser Azure CLI pour gérer des images conteneur et exécuter des tâches rapides ACR

DÉPLOYER DES CONTENEURS SUR AZURE APP SERVICE

- Déployer des conteneurs personnalisés sur Azure App Service à partir de registres de conteneurs.
- Configurer le comportement du runtime de conteneur, notamment les commandes de démarrage, les paramètres de port et le stockage persistant.
- Configurer les paramètres d'application et les chaînes de connexion dans App Service.
- Observer et résoudre les problèmes liés aux applications conteneurisées à l'aide des outils de diagnostic App Service.

DÉPLOYER DES CONTENEURS SUR AZURE CONTAINER APPS

- Expliquer comment les environnements Azure Container Apps affectent la mise en réseau, la journalisation et l'isolation.
- Déployer une application conteneur à l'aide d'Azure CLI et d'Azure CLI avec une définition YAML.
- Configurer les paramètres d'exécution à l'aide de variables d'environnement et de secrets.
- Configurer l'authentification par extraction d'images pour les registres privés.
- Analyser l'intégrité de l'application conteneur à l'aide des journaux, des révisions et de l'état de la réplique.

GÉRER DES CONTENEURS DANS AZURE CONTAINER APPS

- Mettre à jour les images conteneur et gérez les révisions en toute sécurité pendant les cycles de développement et de mise en production.
- Effectuer des opérations de cycle de vie des applications et diagnostiquez les révisions défaillantes.
- Surveiller les journaux de bord et résolvez les problèmes fréquents de révision et d'exécution.
- Configurer les sondes d'intégrité et résolvez les échecs des sondes.
- Optimiser les ressources de conteneur et les paramètres de mise à l'échelle pour équilibrer les coûts et les performances.

METTRE A L'ÉCHELLE DES CONTENEURS DANS AZURE CONTAINER APPS

- Configurer des règles d'échelle HTTP, TCP, PROCESSEUR et mémoire pour les applications conteneur
- Implémenter une mise à l'échelle pilotée par les événements à l'aide de scalers KEDA pour les services Azure
- Sélectionner les ressources de calcul appropriées pour optimiser les performances et les coûts
- Appliquer des modes de révision pour contrôler le comportement de mise à l'échelle et la distribution du trafic

DÉPLOYER DES APPLICATIONS SUR AZURE KUBERNETES SERVICE

- Créer des manifestes de déploiement Kubernetes
- Déployer et exposer des applications sur Azure Kubernetes Service

CONFIGURER DES APPLICATIONS SUR AZURE KUBERNETES SERVICE

- Expliquer comment externaliser la configuration des applications à l'aide de primitives Kubernetes.
- Implémenter ConfigMaps et injectez des paramètres dans des pods.
- Implémenter des secrets et consommez des valeurs sensibles en toute sécurité.
- Attacher un stockage persistant à l'aide de PersistentVolume et persistentVolumeClaim.

SURVEILLER ET DÉPANNER DES APPLICATIONS SUR AZURE KUBERNETES SERVICE

- Expliquer les signaux de surveillance clés pour les charges de travail IA s'exécutant sur Azure Kubernetes Service.
- Utiliser kubectl ainsi que les outils Azure pour examiner les journaux d'activité et les métriques des applications.
- Résoudre des incidents affectant les pods et les services ayant un impact sur les API et les workers d'IA
- Vérifier les chemins de connectivité de service et d'entrée pour vous assurer que les clients peuvent accéder aux services IA.

GÉNÉRER DES REQUÊTES POUR AZURE COSMOS DB POUR NOSQL

- Expliquer le modèle de ressource Azure Cosmos DB pour NoSQL et la façon dont les bases de données, les conteneurs et les éléments se rapportent les uns aux autres
- Implémenter des opérations du Kit de développement logiciel (SDK) pour se connecter à Azure Cosmos DB et effectuer des opérations CRUD sur des éléments

- Sélectionner entre les lectures de points et les requêtes en fonction des exigences de performances et des modèles d'accès
- Créer des requêtes à l'aide de la syntaxe SQL pour filtrer, projeter et récupérer des données à partir de conteneurs

IMPLÉMENTER LA RECHERCHE VECTORIELLE SUR AZURE COSMOS DB POUR NOSQL

- Stocker et récupérer des incorporations de vecteurs dans des conteneurs Azure Cosmos DB avec des stratégies vectorielles correctement configurées
- Exécuter des requêtes de similarité vectorielle à l'aide de la fonction VectorDistance pour rechercher sémantiquement des documents similaires
- Combiner la recherche vectorielle avec des filtres de métadonnées et une recherche en texte intégral à l'aide de requêtes hybrides
- Implémenter le traitement du flux de modification pour actualiser automatiquement les incorporations lorsque les documents sources changent

OPTIMISER LES PERFORMANCES DES REQUÊTES POUR AZURE COSMOS DB POUR NOSQL

- Analyser les modèles de requête et utiliser des métriques pour identifier les goulots d'étranglement des performances et les index manquants
- Configurer des index de plage et composites pour optimiser les opérations de filtrage et de tri pour les modèles de récupération IA
- Sélectionner le type d'index vectoriel approprié en fonction de la taille du jeu de données, des exigences de précision et des objectifs de performances
- Concevoir des stratégies d'indexation qui équilibrent les performances de lecture par rapport aux coûts d'écriture pour votre profil de charge de travail
- Choisir des niveaux de cohérence qui répondent aux exigences de l'application tout en réduisant la consommation de RU

GÉNÉRER ET INTERROGER AVEC AZURE DATABASE POUR POSTGRESQL

- Expliquer l'architecture et les fonctionnalités clés d'Azure Database pour PostgreSQL
- Établir des connexions sécurisées à PostgreSQL à l'aide de l'authentification Microsoft Entra et tls
- Créer et gérer des schémas de base de données, notamment des tables, des index et des contraintes

- Écrire des requêtes SQL efficaces pour les opérations courantes de données
- Intégrer Azure Database pour PostgreSQL dans des applications à l'aide de Python

IMPLÉMENTER LA RECHERCHE VECTORIELLE AVEC AZURE DATABASE POUR POSTGRESQL

- Stocker et interroger des incorporations de vecteurs à l'aide de l'extension pgvector dans Azure Database pour PostgreSQL
- Exécuter des recherches de similarité vectorielle à l'aide de métriques et d'opérateurs de distance différents
- Créer et gérer des index vectoriels pour optimiser les performances de recherche
- Implémenter des stratégies de mise à jour et d'actualisation incorporées pour l'évolution des jeux de données
- Créer des modèles de récupération qui intègrent la recherche vectorielle PostgreSQL avec des pipelines RAG

OPTIMISER LA RECHERCHE VECTORIELLE DANS AZURE DATABASE POUR POSTGRESQL

- Paramétrer les paramètres de configuration PostgreSQL et pgvector pour optimiser la latence des requêtes et l'utilisation de la mémoire pour les charges de travail IA
- Sélectionner et configurer le type d'index vectoriel approprié en fonction de la taille du jeu de données, des modèles de requête et des exigences de précision
- Concevoir des dispositions de données qui optimisent les performances de filtrage des métadonnées et du stockage vectoriel
- Mettre à l'échelle la Base de données Azure pour PostgreSQL afin de gérer les charges de travail vectorielles de grande envergure
- Implémenter des stratégies de regroupement de connexions et de gestion de session pour les applications IA

IMPLÉMENTER DES OPÉRATIONS DE DONNÉES DANS AZURE MANAGED REDIS

- Expliquer les fonctionnalités d'Azure Managed Redis et les stratégies de mise en cache pour les applications hautes performances
- Sélectionner les bibliothèques clientes appropriées et appliquez les meilleures pratiques de développement pour les implémentations Redis
- Implémenter des opérations de données, notamment le

stockage, la récupération, l'expiration et les modèles d'invalidation du cache

IMPLÉMENTER LA MESSAGERIE D'ÉVÉNEMENTS AVEC AZURE MANAGED REDIS

- Expliquer la messagerie publication/abonnement Redis pour diffuser des événements à plusieurs services IA simultanément
- Implémenter des flux Redis pour des files d'attente de tâches fiables avec une nouvelle tentative automatique et une récupération d'échec
- Choisir entre pub/sub et stream selon que vous avez besoin d'une diffusion ou d'une distribution de travail coordonnée
- Créer des applications Python qui utilisent pub/sub pour les notifications et les flux pour le traitement des pipelines

IMPLÉMENTER LE STOCKAGE VECTORIEL DANS AZURE MANAGED REDIS

- Créer des index vectoriels et des incorporations de requêtes pour la recherche de similarité à l'aide de Redis en tant que base de données vectorielle
- Choisir les types de vecteurs appropriés, les métriques de distance et les algorithmes d'indexation en fonction des exigences de taille et de précision du jeu de données
- Sélectionner des structures de données Redis optimales (hachage et JSON) pour stocker des vecteurs avec des métadonnées
- Créer des applications Python qui indexent et interrogent des incorporations à haute dimension avec Azure Managed Redis

METTRE EN FILE D'ATTENTE ET TRAITER DES OPÉRATIONS D'IA AVEC AZURE SERVICE BUS

- Expliquer comment Azure Service Bus découple les composants d'application IA et identifier quand appliquer des modèles de messagerie tels que le nivellement de charge, les consommateurs concurrents et la publication/souscription.
- Choisir entre les files d'attente et les rubriques Service Bus avec des abonnements selon qu'un flux de travail IA nécessite un traitement pour un consommateur unique ou une répartition pour plusieurs consommateurs.
- Structurer les messages Service Bus pour les charges de travail d'IA, notamment la sérialisation des invites et des paramètres de modèle, la gestion de charges utiles volumineuses avec le modèle de réclamation-vérification, ainsi que les ID de corrélation pour le suivi des demandes de bout en bout.
- Traiter les messages de manière fiable à l'aide du mode de

réception peek-lock, gérez les messages incohérents via des files d'attente de lettres mortes et surveillez la file d'attente de lettres mortes pour détecter les échecs d'inférence.

DÉVELOPPER DES FLUX DE TRAVAIL IA PILOTÉS PAR LES ÉVÉNEMENTS AVEC AZURE EVENT GRID

- Expliquer comment Azure Event Grid active les modèles pilotés par les événements dans les solutions IA et identifie les composants principaux (rubriques, abonnements aux événements et gestionnaires d'événements) qui forment une architecture de routage des événements.
- Concevoir des événements à l'aide du schéma CloudEvents pour les opérations IA, définissez des types d'événements personnalisés et configurez des abonnements aux événements avec des filtres qui routent les événements en fonction du type, du sujet ou des attributs de données.
- Configurer des stratégies de remise et de nouvelle tentative pour gérer les défaillances temporaires dans les pipelines IA, définir des destinations de lettres mortes pour les événements non livrables et surveiller les résultats de la livraison.
- Publier des événements personnalisés à partir d'applications IA pour signaler les inférences terminées, les mises à jour de modèle ou les transitions d'étape de pipeline à l'aide du KIT de développement logiciel (SDK) Event Grid et de l'API REST.

CRÉER DES BACK-ENDS D'IA SERVERLESS AVEC AZURE FUNCTIONS

- Évaluer les compromis entre démarrage à froid, mise à l'échelle et mémoire d'instance lors du choix entre l'hébergement Flex Consumption et Premium pour les charges de travail IA
- Configurer un environnement de développement local pour Azure Functions à l'aide de Core Tools, d'émulateurs et d'un IDE
- Créer des déclencheurs et des liaisons qui implémentent des modèles d'intégration IA courants tels que des points de terminaison d'inférence HTTP et des processeurs par lots basés sur une file d'attente
- Configurer la gestion des secrets et les paramètres d'application à l'aide de références Key Vault et d'Azure App Configuration
- Appliquer une identité managée et une autorisation au niveau de la fonction pour sécuriser l'accès entre Functions et d'autres ressources Azure

GÉRER LES SECRETS D'APPLICATION AVEC AZURE KEY VAULT

- Expliquer comment Azure Key Vault stocke et organise les secrets, les clés et les certificats, et identifier quand utiliser

chaque type d'objet dans une solution IA.
- Récupérer les secrets par programmation à l'aide des bibliothèques clientes du Kit de développement logiciel (SDK) Azure avec l'authentification d'identité managée.
- Gérer le contrôle de version et la rotation des versions secrètes dans le code de l'application pour prendre en charge les mises à jour des informations d'identification sans temps d'arrêt.
- Implémenter des stratégies de mise en cache qui réduisent les appels d'API Key Vault tout en conservant les garanties de sécurité et de fraîcheur.

GÉRER LES PARAMÈTRES D'APPLICATION AVEC AZURE APP CONFIGURATION

- Se connecter à Azure App Configuration à partir du code de l'application et récupérer les paramètres à l'aide de la bibliothèque de fournisseurs Python avec l'authentification d'identité managée.
- Organiser les paramètres de configuration avec des étiquettes et implémentez des indicateurs de fonctionnalité pour contrôler la disponibilité des fonctionnalités sans redéployer.
- Référencer les secrets Azure Key Vault depuis App Configuration pour unifier l'accès à la configuration et aux secrets dans un seul et même chemin de récupération.
- Déterminer les paramètres qui appartiennent à App Configuration et qui appartiennent à Key Vault en fonction de la sensibilité, de la structure et des modèles d'accès.

INSTRUMENTER UNE APPLICATION AVEC OPENTELEMETRY

- Expliquer comment OpenTelemetry fournit une observabilité neutre par le fournisseur pour les applications IA distribuées sur Azure.
- Ajouter et configurer la distribution OpenTelemetry Azure Monitor dans une application pour collecter des données de télémétrie.
- Créer et gérer des étendues et des traces personnalisées pour capturer les flux de demandes entre les services distribués.
- Exporter les données de télémétrie vers Azure Monitor Application Insights pour l'analyse et la visualisation.
- Utiliser des données de trace dans Application Insights pour identifier et déboguer les problèmes de performances dans les flux de travail distribués.

ANALYSER LES DONNÉES DE TÉLÉMÉTRIE DES APPLICATIONS AVEC DES JOURNAUX ET DES MÉTRIQUES

- Écrire des requêtes KQL pour récupérer et analyser les données de télémétrie d'application à partir d'Application Insights.
- Explorer les données de journal pour identifier les modèles d'erreur, les goulots d'étranglement des performances et les tendances dans le comportement de l'application.
- Créer des tableaux de bord Azure qui affichent les métriques de télémétrie clés et les résultats des requêtes de journal pour la surveillance opérationnelle.
- Créer des classeurs Azure Monitor pour l'analyse interactive des données de télémétrie pilotées par les paramètres.

Accessibilité

L'inclusion est sujet important pour OCTO Academy.
Nos référent·es sont à votre disposition pour faciliter l'adaptation de votre formation à vos besoins spécifiques.
Pour les contacter : academy.accessibilite@octo.com