



Laboratoire pratique : IA générative pour la configuration de l'infrastructure

Effort estimé : 30 minutes

Introduction

Dans le monde des entreprises modernes axées sur les données, le choix de l'infrastructure de données constitue une décision cruciale, qui influence non seulement l'efficacité de la gestion des données, mais aussi la capacité d'innovation et l'avantage concurrentiel de l'organisation. Dans ce laboratoire pratique, nous nous penchons sur la synergie transformatrice entre GenAI et l'infrastructure de données, en expliquant comment les technologies GenAI peuvent révolutionner les pratiques traditionnelles d'ingénierie des données.

Il est primordial de comprendre l'importance de sélectionner la bonne infrastructure de données pour s'y retrouver dans les complexités des écosystèmes de données contemporains. De plus, l'exploitation de la puissance de l'IA générative apporte de nouvelles solutions aux défis de longue date, favorisant l'agilité, l'évolutivité et l'adaptabilité au sein des opérations de données. Dans ce laboratoire, vous explorerez divers scénarios dans lesquels GenAI croise l'infrastructure de données, présentant des défis du monde réel et sollicitant des suggestions pour des configurations d'infrastructure optimales.

Objectif(s)

À la fin de ce laboratoire, vous serez en mesure d'utiliser l'IA générative pour proposer des exigences d'infrastructure de données pour différents scénarios pratiques dans différents secteurs couvrant les scénarios suivants :

- Commerce de détail
- Secteur de la santé
- Secteur financier
- Industrie des médias et du divertissement

Commerce de détail

Scénario

Une plateforme de commerce électronique anticipe une augmentation significative des commandes pendant la période des fêtes à venir. Son infrastructure de données actuelle a du mal à gérer l'augmentation attendue du volume, de la vitesse et de la variété des données. Par conséquent, les ingénieurs de données de l'entreprise proposent d'améliorer leur infrastructure de données. Des discussions sont en cours sur les aspects du stockage évolutif, de meilleures capacités de traitement et d'analyse en temps réel.

Rapide

Pour utiliser l'IA générative pour des entrées suggestives sur cet énoncé de problème, vous pouvez utiliser l'invite suivante.

```
How should an E-commerce platform enhance their data infrastructure, such that they are able to handle increase in traffic. Suggest  
1. scalable storage,  
2. better processing capabilities and  
3. real-time analytics.
```

La solution générée peut avoir une formulation différente à chaque fois, mais globalement la structure de la réponse restera la même. L'une des façons dont vous pouvez trouver la réponse des systèmes GenAI est partagée ci-dessous.

Pour améliorer leur infrastructure de données afin de gérer une augmentation du trafic, une plateforme de commerce électronique peut envisager les améliorations suivantes :

1. Stockage évolutif : la mise en œuvre d'une solution de stockage basée sur le cloud, telle qu'Amazon S3 ou Google Cloud Storage, peut fournir l'évolutivité nécessaire pour gérer une augmentation du volume de données. Ces services offrent une mise à l'échelle automatique en fonction de la demande, garantissant que la plateforme peut gérer un trafic accru sans compromettre les performances.
2. Meilleures capacités de traitement : la mise à niveau vers une infrastructure de serveur plus puissante ou l'utilisation d'un framework de calcul distribué comme Apache Hadoop ou Spark peuvent améliorer les capacités de traitement. Ces technologies permettent un traitement parallèle des données, ce qui permet un traitement plus rapide et plus efficace des données, ce qui est essentiel pour gérer de gros volumes de trafic.
3. Analyse en temps réel : la mise en œuvre d'une plateforme d'analyse en temps réel, telle qu'Apache Kafka ou Apache Flink, peut permettre d'analyser les données au fur et à mesure de leur arrivée, ce qui permet d'obtenir des informations et des prises de décision immédiates. L'analyse en temps réel peut aider la plateforme de commerce électronique à optimiser l'expérience utilisateur, à personnaliser les recommandations et à détecter les fraudes ou les anomalies en temps réel.

En intégrant ces améliorations en matière de stockage évolutif, de meilleures capacités de traitement et d'analyses en temps réel, une plateforme de commerce électronique peut améliorer son infrastructure de données pour gérer efficacement une augmentation du trafic et offrir une expérience utilisateur transparente.

Secteur de la santé

Scénario

Un prestataire de soins de santé souhaite établir un référentiel de données central à des fins de recherche. Ce lac de données contiendrait des données sur les patients, des dossiers médicaux et des études de recherche. Les ingénieurs de données de l'entreprise discutent de la mise en œuvre d'outils de gestion du Big Data, de capacités d'ingestion et de transformation des données et de directives de sécurité et de conformité des données.

Rapide

Pour utiliser l'IA générative pour des entrées suggestives sur cet énoncé de problème, vous pouvez utiliser l'invite suivante :

How should a healthcare company set up a data lake infrastructure that is capable of the following.

1. Big data management
2. data ingestion from various sources
3. data transformation
4. data security and compliance to regulatory guidelines

La solution générée peut avoir une formulation différente à chaque fois, mais globalement, la structure de la réponse restera la même. L'une des façons dont vous pouvez trouver la réponse des systèmes GenAI est partagée ci-dessous.

Pour mettre en place une infrastructure de lac de données pour une entreprise de soins de santé qui réponde à ces exigences, les étapes suivantes peuvent être suivies :

1. Gestion du Big Data : utilisez des solutions de stockage cloud évolutives telles qu'Amazon S3 ou Google Cloud Storage pour stocker de gros volumes de données. Mettez en œuvre des techniques de partitionnement et d'indexation des données pour gérer et interroger efficacement les données.
2. Ingestion de données : utilisez des outils comme Apache NiFi ou Apache Kafka pour ingérer des données provenant de diverses sources telles que des dossiers médicaux électroniques, des appareils médicaux et des objets connectés. Assurez-vous que le processus d'ingestion de données est fiable, évolutif et capable de gérer des données en streaming en temps réel.
3. Transformation des données : implémentez des frameworks de traitement de données comme Apache Spark ou Apache Flink pour transformer les données brutes en un format structuré adapté à l'analyse. Utilisez des pipelines de données et des processus ETL pour nettoyer, enrichir et agréger les données avant de les stocker dans le lac de données.
4. Sécurité et conformité des données : implémentez des techniques de chiffrement pour sécuriser les données au repos et en transit. Utilisez des mécanismes de contrôle d'accès pour restreindre l'accès aux données en fonction des rôles et des autorisations des utilisateurs. Assurez la conformité aux réglementations telles que HIPAA et GDPR en mettant en œuvre des pistes d'audit, le masquage des données et des techniques d'anonymisation.

En suivant ces étapes, une entreprise de soins de santé peut établir une infrastructure de lac de données robuste qui répond aux exigences de gestion du Big Data, d'ingestion de données, de transformation des données, de sécurité et de conformité des données.

Secteur financier

Scénario

Une société de services financiers souhaite mettre en œuvre des modèles d'apprentissage automatique pour détecter les transactions frauduleuses en temps réel. Cela nécessite de créer et de maintenir une infrastructure de données robuste pour prendre en charge les modèles. Cela nécessite que les ingénieurs de données mettent en place des machines de calcul haute performance, des pipelines d'ingénierie des fonctionnalités, une modélisation prédictive, ainsi que le déploiement et la surveillance des modèles.

Rapide

Pour utiliser l'IA générative pour des entrées suggestives sur cet énoncé de problème, vous pouvez utiliser l'invite suivante.

How should a financial firm set up their infrastructure if they want to detect fraudulent transactions in real time? Suggest specific

1. Computing machinery
2. Feature engineering pipeline
3. Predictive modeling pipeline
4. Model deployment and monitoring

La solution générée peut avoir une formulation différente à chaque fois, mais globalement, la structure de la réponse restera la même. L'une des façons dont vous pouvez trouver la réponse des systèmes GenAI est partagée ci-dessous.

1. Machines informatiques : pour les machines informatiques, une société financière pourrait envisager d'investir dans des serveurs hautes performances dotés de processeurs multicœurs et d'une mémoire suffisante pour gérer le traitement de grands ensembles de données en temps réel. Elle pourrait également explorer les services de cloud computing tels qu'AWS, Google Cloud ou Microsoft Azure pour l'évolutivité et la flexibilité dans la gestion des ressources informatiques.
2. Pipeline d'ingénierie des fonctionnalités : en termes de pipeline d'ingénierie des fonctionnalités, l'entreprise financière pourrait utiliser des outils tels que les bibliothèques Python (par exemple, Pandas, NumPy) pour la manipulation des données et l'extraction des fonctionnalités. Elle pourrait également exploiter des techniques de sélection de fonctionnalités telles que l'élimination récursive des fonctionnalités ou l'analyse des composants principaux pour identifier les fonctionnalités les plus pertinentes pour la détection des fraudes.
3. Pipeline de modélisation prédictive : pour le pipeline de modélisation prédictive, l'entreprise financière pourrait envisager d'utiliser des cadres d'apprentissage automatique tels que TensorFlow ou scikit-learn pour créer et former ses modèles. Elle pourrait également explorer des techniques telles que la validation croisée et le réglage des hyperparamètres pour optimiser les performances de ses modèles afin de détecter les transactions frauduleuses.
4. Déploiement et surveillance des modèles : En termes de déploiement et de surveillance des modèles, l'entreprise financière pourrait utiliser des outils de conteneurisation comme Docker pour emballer ses modèles et les déployer dans un environnement de production. Elle pourrait également mettre en œuvre des outils de surveillance comme Prometheus ou Grafana pour suivre les performances de ses modèles en temps réel et configurer des alertes en cas d'anomalie ou d'écart dans la précision de la détection. Le recyclage et la mise à jour réguliers des modèles pourraient être automatisés à l'aide d'outils comme Jenkins ou Airflow pour garantir l'efficacité continue du système de détection des fraudes.

Pratique - Industrie des médias et du divertissement

Scénario

Une entreprise de médias et de divertissement souhaite personnaliser l'expérience utilisateur en recommandant du contenu en fonction des préférences individuelles. Cela nécessite de créer et de maintenir une infrastructure de données complexe pour prendre en charge le moteur de recommandation. Les ingénieurs de données discutent donc des défis infrastructurels en termes d'ingestion de données en temps réel, d'entreposage de données, de cadres de traitement parallèle et de développement de modèles d'apprentissage automatique pour cette tâche.

Rapide

Nous vous encourageons à créer une invite pour ce scénario et à générer les exigences d'infrastructure pour la tâche.

Conclusion

Félicitations pour avoir terminé ce laboratoire.

Vous êtes désormais en mesure d'utiliser l'IA générative pour créer des exigences d'infrastructure pour différents secteurs en fonction de différents scénarios dans chacun d'eux.

Auteur(s)

[Abhishek Gagneja](#)

© IBM Corporation. Tous droits réservés.