

Cloud environment (GCP: Google Cloud Platform)

1. What is the best way to migrate your application to be deployed on the AI services of GCP?

Après avoir entraîné le modèle par exemple un Modèle Tensorflow dans **Google Storage, bucket Cloud Storage**, le modèle doit être hébergé dans le **AI Platform**. Ensuite, il faut **bâtir l'application App**, par exemple avec Streamlit. Ensuite, il faut **bâtir un Container Docker** avec cette application. Par la suite, le Container Docker va être déployé sur **Google Container Registry (GCR)**. Ensuite, le container hébergé dans GCR va être **déployé dans App Engine**. Et enfin, l'application va être **monitorée avec Google Monitoring**.

2. What modification do you propose to your code to make your training jobs compatible with the training pipelines services on GCP?

Tout d'abord, avant de pouvoir envoyer une tâche d'entraînement, il faut emballer l'application et l'importer avec toutes les dépendances non standards, dans un bucket Cloud Storage. Ensuite, au niveau des services pipelines d'entraînement, il est crucial de configurer le clusters avec un niveau d'évolutivité (exemple : CUSTOM), indiquant le type de cluster de traitement sur lequel la tâche va s'exécuter.

3. What is the best way to manage your datasets and model versioning on GCP?

Il est judicieux de stocker les données dans le projet Google Cloud où le modèle est stocké afin que le projet AI ait accès à tous les ensembles de données requis pour la modélisation. Mais dans la réalité, cela n'est pas tout le temps possible. Maintenant, puisqu'on a différentes versions des modèles, il est aussi important d'avoir les versions de données associées à la bonne version du modèle. Pour se faire, il faut utiliser Data Version Control (DVC) avec Google Cloud Storage et Python pour les modèles. Justement, pour la gestion des versions des modèles, on peut utiliser Vertex AI Model Registry. Avec Vertex AI Model Registry, on peut afficher les modèles et toutes leurs versions dans une même vue. On peut afficher le détail des versions de modèles spécifiques et voir avec précision leurs performances.