

Into the cloud... based on a true story

Comment une DeepTech entraîne ses modèles d'IA chez OVHcloud





24 000 catégories
de produits



jusqu'à 2,5 To
de données
d'entraînement et de
modèles de machine
learning



Base de données
de 250 000 lignes

Le contexte

Créée en Octobre 2020, Customs Bridge est une «DeepTech», une start-up dont le cœur de la technologie s'appuie sur des algorithmes d'intelligence artificielle afin de créer un moteur de classification de produits automatique. La start-up destine ce service aux importateurs européens car chaque produit importé dans l'Union Européenne doit être précisément catégorisé selon une nomenclature qui compte plus de 24 000 entrées. Toute la complexité pour les importateurs est de choisir la bonne catégorie à partir du descriptif fourni par le fabricant, un descriptif qui peut être très succinct voire incomplet. Tout produit importé dans l'Union Européenne doit être déclaré selon un code en fonction duquel les droits de douane sont calculés. Ce code définit aussi la réglementation qui va s'appliquer au produit. Toute erreur d'affectation peut donner lieu à des sanctions, à un retrait du marché et à des redressements fiscaux.

«Nous avons pu bénéficier du programme «Start-Up» d'OVHcloud, ce qui nous a permis de commencer à utiliser leurs services Cloud dédiés à l'IA très rapidement. Le service OVHcloud AI Training nous a permis ainsi de réaliser l'apprentissage de nos modèles de Machine Learning qu'il nous était impossible de mener sur nos propres machines on-premise.»

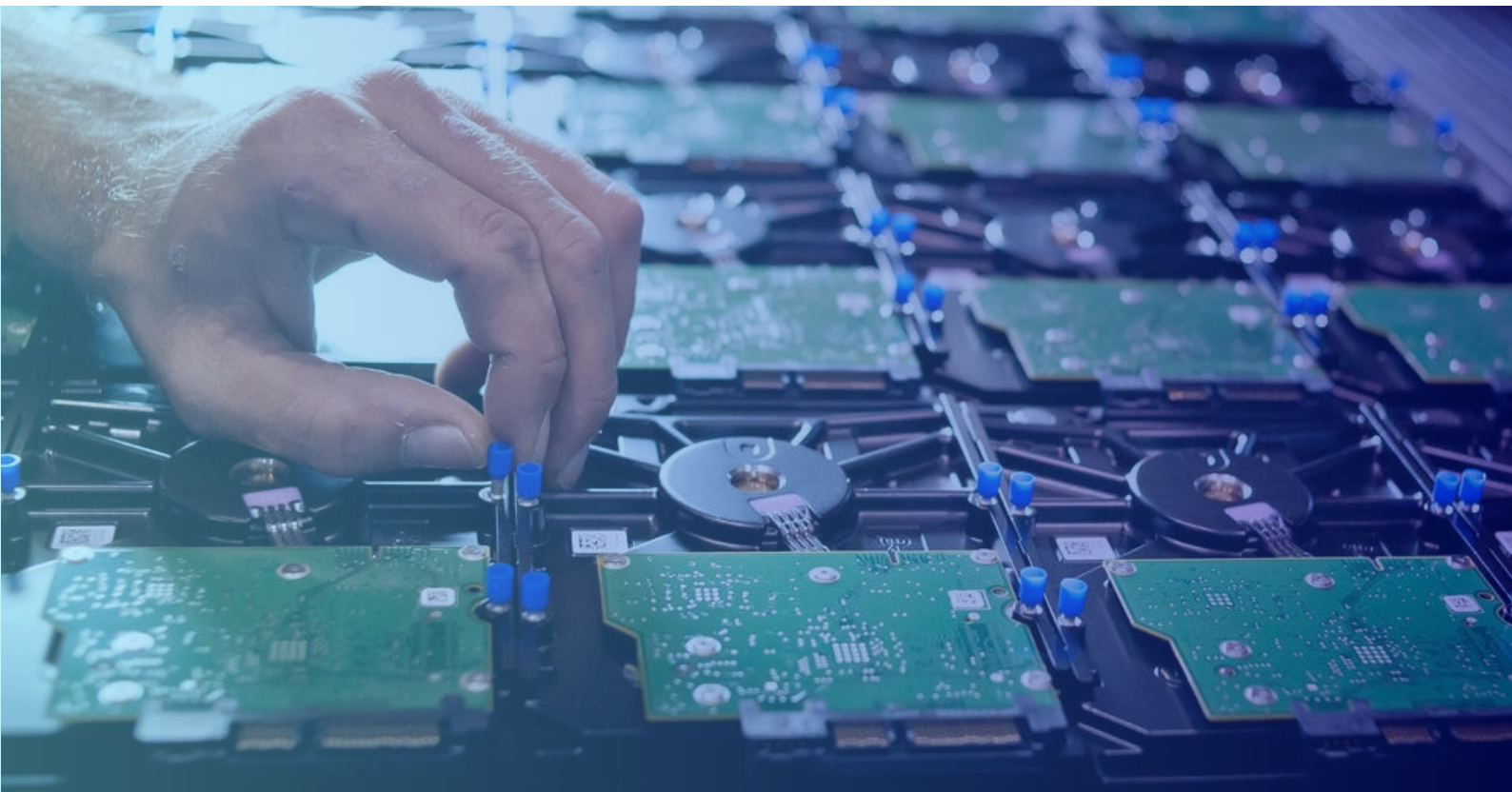
Dr. Hamza Saouli, directeur de l'innovation de Customs Bridge

Cette classification peut s'avérer problématique car ce code doit être uniforme dans l'ensemble des pays de l'Union Européenne mais à l'international les catégories peuvent varier d'un pays à l'autre, selon que le fabricant exporte ses produits en Europe, aux Etats-Unis ou en Chine. Des subtilités dans le descriptif d'un produit peuvent aussi le faire basculer d'une catégorie à une autre, un bracelet-montre n'étant pas classifié de la même façon qu'une chaîne de montre...

Le défi

L'objectif de Customs Bridge est de créer un moteur de classification de produits le plus fiable possible afin d'affecter le bon code douanier à un produit dont le descriptif n'est absolument pas formalisé : il peut s'agir d'un descriptif relativement précis dans le cas des produits électroniques, par exemple, ou quelques mots-clés pour un produit alimentaire, avec des volumes de données très différents selon qu'il s'agit d'un produit fréquemment importé dans l'Union européenne ou pas.

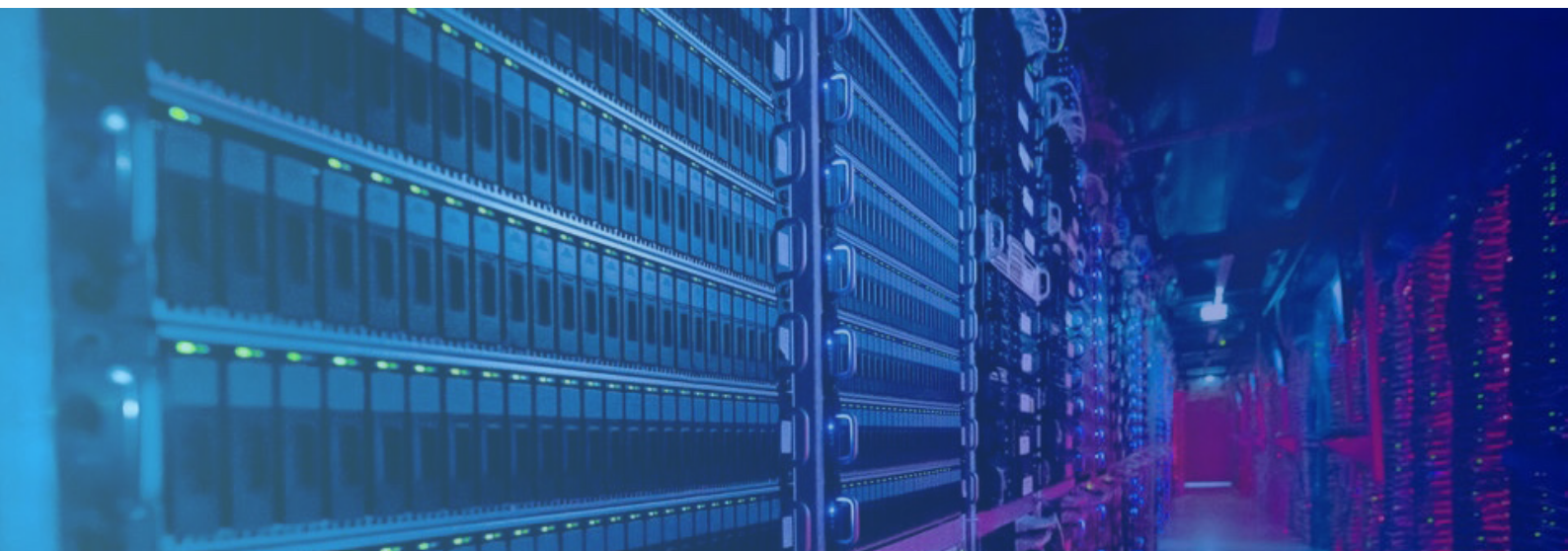
«Pour réaliser l'apprentissage de nos modèles d'intelligence artificielle, nous avons commencé par utiliser des données accessibles en Open Data, notamment la base de données communautaire EBTI (European Binding Tariff Information)» raconte Hamza Saouli, directeur de l'innovation de Customs Bridge. *«Cette base de données compte 250 000 lignes mais elle ne couvre que 10% à 15% de la nomenclature complète. Nous avons pu lancer l'entraînement de plusieurs modèles d'apprentissage sur cette source de données avec des premiers résultats probants sur un code, un chapitre. Sur les produits électroniques importés de Chine, généralement bien décrits, cet apprentissage a obtenu de bons résultats mais sur des produits moins fréquemment importés nous n'avons pas eu de résultats probants faute de données disponibles en grande quantité et de bonne qualité.»* Les modèles ne disposent souvent pas assez de données sur les produits rarement importés sachant que les données européennes sont beaucoup moins accessibles que celles des douanes américaines, par exemple.



Dans les phases initiales du projet, le directeur de l'innovation de Customs Bridge a principalement utilisé les algorithmes IA les plus connus pour leur efficacité et leur rapidité telles que le SVM et les arbres de décision, mais avec l'augmentation de la taille du jeu de données d'entraînement, l'utilisation de ces derniers s'est avéré ne plus être une bonne solution, ce qui a poussé l'équipe IA de Customs Bridge à adopter des modèles plus avancés tels que les réseaux de neurones (via l'API de Deep Learning Keras) et les Transformers, des algorithmes qui sont aujourd'hui à l'état de l'art dans la classification sémantique. L'expert s'est ensuite appuyé sur des articles scientifiques de chercheurs en IA afin d'accroître les performances de ses modèles en matière de classification. Dès lors, la start-up s'est rapidement heurtée à un problème de taille : la capacité de traitement disponible pour entraîner ses modèles d'IA. Si les 3 PC équipés de GPU étaient suffisants pour entraîner les modèles les plus simples, cette infrastructure va rapidement arriver à ses limites et pousser l'équipe de Customs Bridge à opter pour une solution Cloud, idéale pour faire face à un besoin d'une forte puissance de calcul et de mémoire RAM de manière intermittente. C'est la raison qui a poussé Customs Bridge à s'intéresser aux offres IA & Machine Learning d'OVHcloud.

«Initialement, nous avons pensé pouvoir réaliser l'entraînement de nos modèles sur nos propres machines dotées de GPU. Cette approche est rapidement arrivée dans une impasse lorsque nous avons souhaité monter en puissance. Nous étions bloqués par le manque de RAM, par l'espace de stockage disponible, ce qui limitait fortement l'apprentissage de nos modèles. Le Cloud était pour nous la meilleure solution possible tant techniquement qu'économiquement.»

Dr. Hamza Saouli, directeur de l'innovation de Customs Bridge



La solution

Parmi l'ensemble des briques fonctionnelles proposées par OVHcloud dans son offre AI, Customs Bridge met en œuvre la solution OVHcloud dédiée à l'entraînement des modèles, AI Training. En parallèle, la start-up exploite des instances OVH afin de déployer ses modèles en production et supporter le pipeline d'alimentation en données. «Nous avons mis en place un pipeline qui part de la demande d'un client, soumet la requête au modèle puis traite la réponse reçue de la part du modèle» explique Hamza Saouli. «Celle-ci doit être préparée avant d'être affichée pour le client. En entrée nous devons donc traiter des descriptions textuelles de produits à importer, sachant que ces derniers sont de petite taille (3 à 5 mots seulement) et ne décrivent pas assez le produit, et en sortie ces descriptions sont téléversées sur le Cloud afin d'être soumises au modèle déployé qui propose un ensemble de codes douaniers pour l'importateur.»

Dans un avenir proche, ce pipeline est appelé à devenir plus complexe. L'équipe est en train de travailler sur un «augmentateur de texte», un algorithme qui part d'un jeu de données existant et qui va l'enrichir afin d'optimiser l'apprentissage des modèles. L'algorithme va ainsi faire passer la base de données initiale de 200 000 / 300 000 lignes pour la faire monter à 3 à 4 millions de lignes via des techniques de génération de texte automatique. Là encore le Cloud est irremplaçable pour une telle tâche car l'entraînement de modèles sur de tels volumes de données n'est tout simplement plus possible sur des PC classiques.

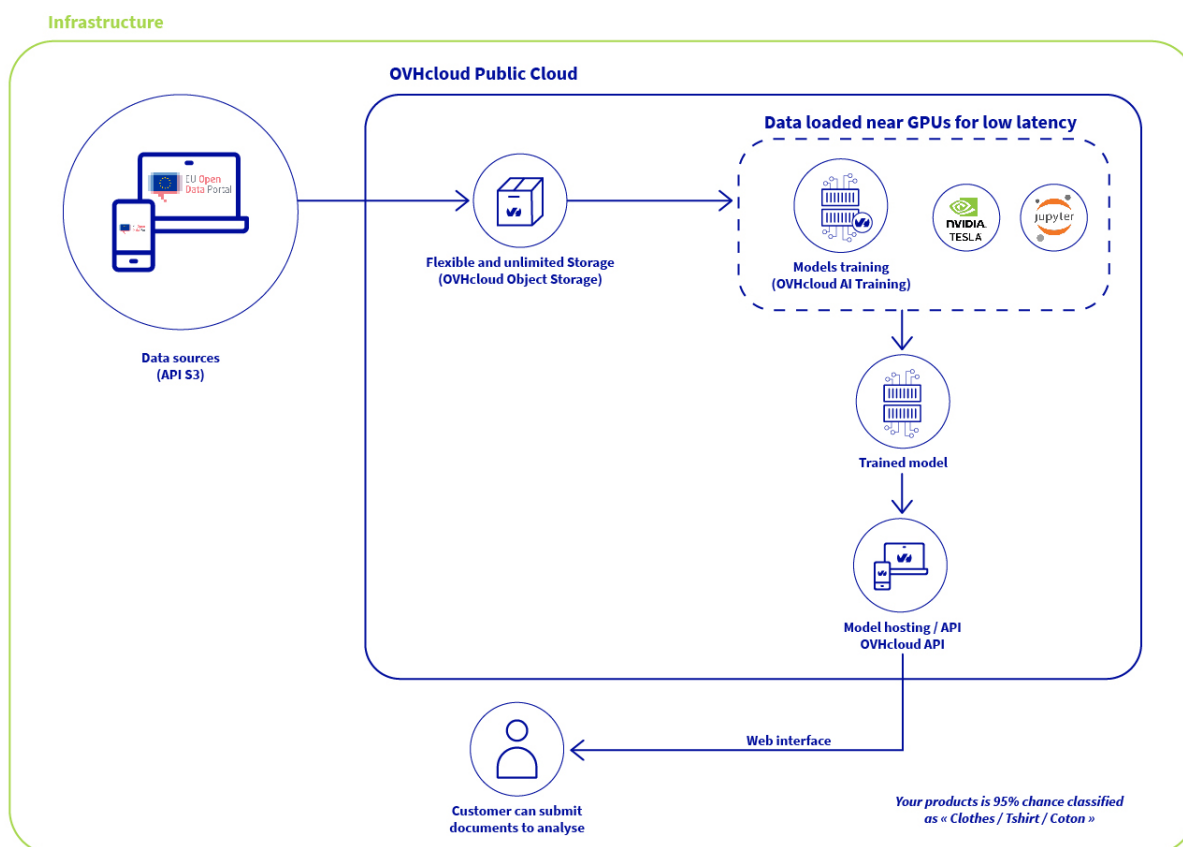
«Basculer l'apprentissage des modèles d'IA d'une approche On-Premise à OVHcloud AI Training nous a apporté une flexibilité et une puissance que nous ne pouvions pas avoir en interne. La solution est très simple à utiliser : Nous pouvons fixer à l'avance le nombre de GPU et la taille de la RAM dont nous allons avoir besoin à l'instant t pour mener un apprentissage. C'est très utile dès lors que l'on connaît à l'avance le nombre de ressources dont on va avoir besoin.»

Dr. Hamza Saouli, directeur de l'innovation de Customs Bridge

Comme l'explique Hamza Saouli, le passage de calcul d'apprentissage en mode on-premise sur le Cloud OVHcloud ne lui a posé aucun problème d'adaptation. OVHcloud fournit des conteneurs prêts à l'emploi pour les principaux frameworks d'IA et il suffit de lancer le Job correspondant pour pouvoir les déployer sur un GPU dans le Cloud. En outre, depuis le mois de juin 2021 il est maintenant possible de faire de même pour les conteneurs exécutés sur CPU. Cette possibilité de choix permet de bénéficier de ressources de calcul à un tarif encore plus bas pour des apprentissages qui ne nécessitent pas la puissance d'un GPU dédié. Cette évolution de l'offre «AI» d'OVHcloud est issue d'une demande de Customs Bridge.

Pour entraîner ses premiers modèles de Transformers, le directeur de l'innovation s'est appuyé sur environ 2,5 To de données. Pour les modèles de Machine Learning, les volumes de données sont plus faibles, de l'ordre de 30 à 40 Go de données d'entraînement. «Avec les GPU NVidia V100 mis à disposition par OVHcloud, l'apprentissage d'un Transformers sur 250 000 lignes ne représente qu'une trentaine de minutes de calcul. C'est à la fois très rapide et le coût est véritablement modique puisqu'une heure de calcul est facturée environ 1,75 €. C'est la raison pour laquelle nous n'envisageons absolument pas d'acquérir de machines pour réaliser ces calculs en interne» ajoute l'expert.

En parallèle à ce travail mené sur les modèles d'IA, Hamza Saouli travaille aujourd'hui sur un chatbot qui interagira avec les clients pour obtenir les informations sur le produit recherché. Celui-ci a déjà entraîné un modèle RASA, une plateforme Open Source dédiée aux chatbots sur des instances CPU d'OVHcloud. Les premiers résultats ont été jugés très encourageants et l'expert espère qu'OVHcloud mettra rapidement à disposition un conteneur RASA dans son infrastructure AI pour en simplifier encore la mise en œuvre.



Le résultat

«Après plusieurs mois d'utilisation d'OVHcloud AI Training et l'entraînement de multiples types de modèles d'IA, je n'ai jamais connu de problèmes d'installation ou de paramétrage» se félicite Hamza Saouli. «OVHcloud nous donne la possibilité de choisir l'image Docker sur laquelle va être lancé l'apprentissage, c'est une approche extrêmement simple et efficace. J'ai utilisé ces conteneurs pour des modèles de Transformers et Tensorflow pour un ChatBot avec les images disponibles et cela fonctionne parfaitement.»

Outre l'apprentissage de ses modèles, phase traditionnellement très coûteuse en espace mémoire et en puissance de calcul, Customs Bridge réfléchit aujourd'hui à la scalabilité de son modèle au niveau de la production lorsque la start-up signera ses premiers clients. «Pour le moment, notre modèle le plus performant est un modèle classique qui ne nécessite pas de GPU pour être déployé en production» explique Hamza Saouli. «Lorsque nous utiliserons des jeux de données plus importants, nous allons augmenter les volumes de données d'un facteur x100 à x1000 dans un avenir proche. Ce facteur nous importe peu et tout dépendra de la pertinence du modèle. C'est tout l'intérêt d'une approche Cloud : OVHcloud nous permettra de faire croître les volumes de données sans contrainte d'infrastructure. Nous n'avons pas à brider nos modèles, nous allons simplement expérimenter jusqu'à évaluer le volume nécessaire pour atteindre la précision que nous recherchons. C'est la liberté que nous apporte le modèle Cloud.»

Customs Bridge utilisera alors des instances GPU en production si le besoin s'en fait sentir. La start-up aura alors la possibilité d'exécuter ses modèles d'IA sur le service OVHcloud ML Serving. «De même, le service Data Preparation d'OVHcloud est potentiellement intéressant pour nous lorsque nous aurons de plus gros volumes de données à traiter en amont de nos modèles. L'allocation dynamique des ressources nous permet de ne payer que ce que l'on consomme réellement et c'est un atout pour Customs Bridge conclut l'expert en intelligence artificielle.

OVHcloud est un acteur mondial et le leader européen du Cloud opérant plus de 400 000 serveurs dans ses 32 datacenters sur 4 continents. Depuis 20 ans, le Groupe s'appuie sur un modèle intégré qui lui confère la maîtrise complète de sa chaîne de valeur : de la conception de ses serveurs à la construction et au pilotage de ces datacenters en passant par l'orchestration de son réseau de fibre optique. Cette approche unique lui permet de couvrir en toute indépendance l'ensemble des usages de ses 1,6 million de clients dans plus de 140 pays. OVHcloud propose aujourd'hui des solutions de dernière génération alliant performance, prédictibilité des prix et une totale souveraineté sur leurs données pour accompagner leur croissance en toute liberté.