

MOBIDEV - CAS D'UTILISATION N°1 DE KUBERNETES : INGÉNIERIE DES DONNÉES POUR UN SYSTÈME DE PRÉVISION DE LA DEMANDE BASÉ SUR L'IA

DESCRIPTION DU PROJET

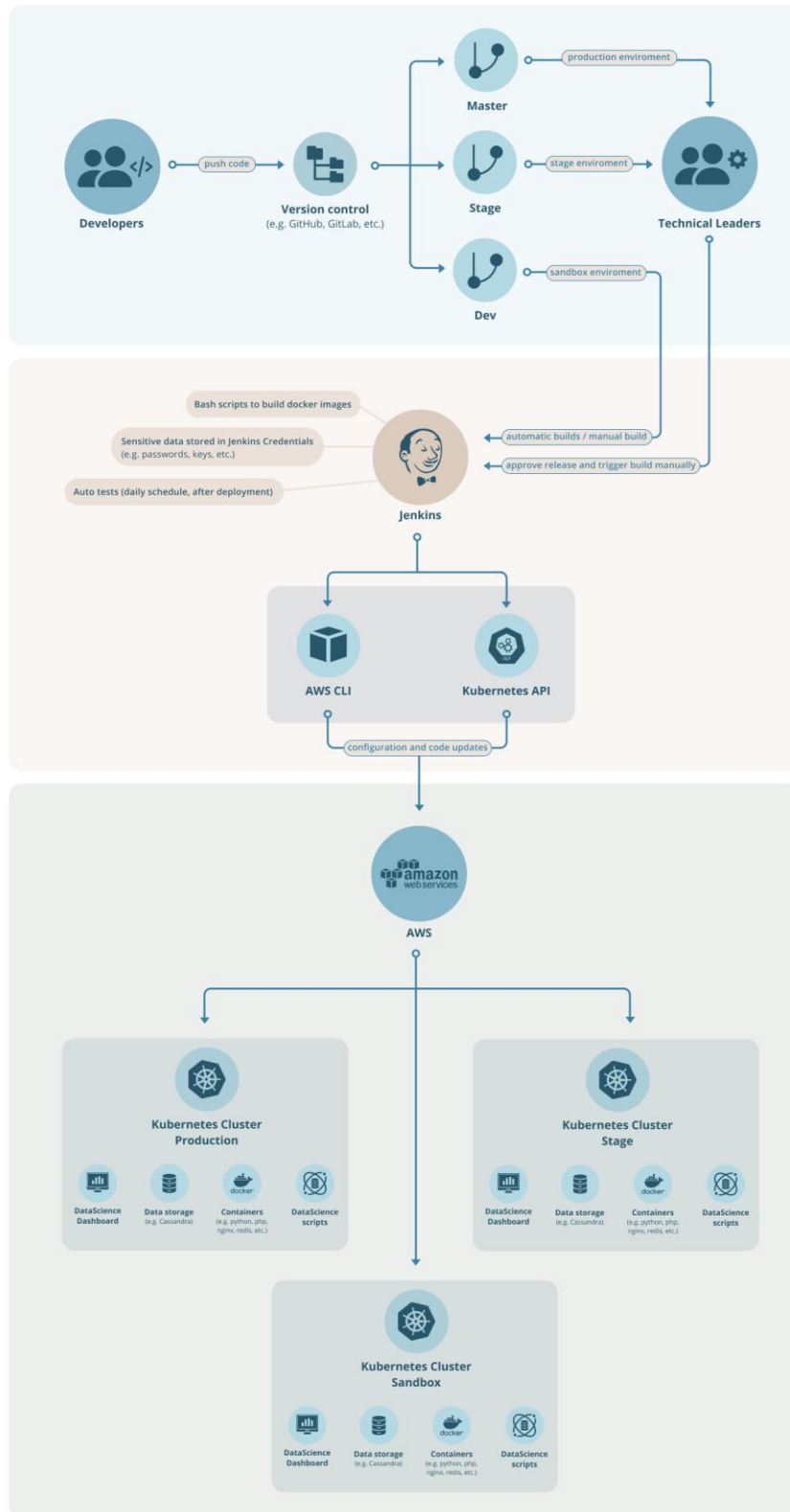
MobiDev avait développé un logiciel de point de vente et un [système de](#) gestion de sites adoptés par de nombreux bars et restaurants. Au fil du temps, une énorme quantité de données historiques sur les ventes de points de vente se sont accumulées et les experts techniques de MobiDev ont eu l'idée d'appliquer des algorithmes d'intelligence artificielle (IA) aux données, de trouver des modèles dans les ventes précédentes et de faire des prédictions pour la prochaine période pour chaque lieu. Un système de [prévision de la demande](#) basé sur l'IA a été développé, en tant que module indépendant intégré au système.

LES PROBLÈMES À RÉSOUDRE

Comme les calculs de l'IA nécessitaient des ressources importantes, au départ, les machines virtuelles du service cloud AWS EMR ont été utilisées. Plus il y avait de sites qui adoptaient le système, plus l'infrastructure devenait chère. La charge CPU élevée du module AI s'est produite pendant la nuit, lorsque les algorithmes d'apprentissage automatique traitaient les données de vente quotidiennes et étaient inactifs pendant la journée. Afin de diminuer les coûts d'infrastructure, une gestion manuelle des ressources informatiques a été mise en place avec l'aide de Docker Swarm.

Au stade du développement du MVP, cela n'avait pas de sens d'utiliser Kubernetes car cela nécessiterait beaucoup de temps de recherche et d'adoption. Cependant, comme le nombre de sites était censé augmenter, une nouvelle approche de l'ingénierie des données était nécessaire pour fournir l'automatisation et l'évolutivité ainsi qu'une optimisation des coûts.

Kubernetes use case: Data engineering for AI-based demand forecasting system



TÂCHES TECHNIQUES RÉSOLUES AVEC KUBERNETES

1. Scripts de collecte de données historiques sur une base planifiée
2. Stockage de données dans une base de données qui s'exécute dans Kubernetes
3. Scripts d'IA après une mise à jour réussie des données historiques
4. API qui interagit avec le tableau de bord AI
5. Tableau de bord AI qui affiche les résultats des scripts AI

TÂCHES COMMERCIALES RÉSOLUES AVEC KUBERNETES

Kubernetes nous a permis de mettre en œuvre l'auto-scaling et a fourni une optimisation des ressources informatiques en temps réel.

Optimisation des performances et des coûts

Il a été remarqué que sur Kubernetes, les scripts d'IA, qui calculaient la même logique pour la même quantité de données, renvoyaient des résultats beaucoup plus rapidement que sur AWS EMR, tout en consommant moins de ressources informatiques qu'EMR. Sur Kubernetes, le temps moyen requis pour exécuter les scripts du module AI a diminué de 10 fois pour le même nombre de sites, par rapport à l'environnement de production EMR précédent.

Amélioration de la fiabilité

La stabilité du système était la principale raison qui nous a poussés à passer d'AWS EMR à Kubernetes. Sur EMR, le démarrage du script échouait parfois pour des raisons inconnues et les journaux ne donnaient aucune information utile.

Amélioration de l'évolutivité

Sur Amazon EMR, le développement du projet a été limité par un nombre maximum de nouveaux sites à ajouter à l'avenir. Kubernetes a supprimé la limitation et la mise à l'échelle automatisée, ce qui est essentiel pour les projets à croissance rapide.

RÉSUMÉ DU PROJET

Le système sur Kubernetes fournit des résultats plus rapidement, consomme moins de ressources informatiques, permet au client de réduire les coûts de facturation AWS et assure une livraison de produit stable et prévisible.

MOBIDEV - CAS D'UTILISATION N°2 DE KUBERNETES : INGÉNIERIE DES DONNÉES POUR UN SYSTÈME DE VIDÉOSURVEILLANCE IA

Un autre cas d'utilisation commerciale réel pour Kubernetes MobiDev a été impliqué, est une mise à l'échelle automatique des ressources informatiques intelligentes pour une fonction de flou de visage dans un système de vidéosurveillance. Le système se compose des applications suivantes : front-end, back-end, files d'attente back-end et une [fonctionnalité basée sur l'IA pour le floutage des visages](#) . Kubernetes est utilisé comme orchestrateur pour toutes ces applications.

Lorsque de nouvelles demandes de traitement vidéo apparaissent, le back-end s'adapte automatiquement à l'aide de l'API Kubernetes et ajoute automatiquement plus de travailleurs pour traiter les demandes.

Références :

<https://mobidev.biz/blog/>