

Conception pour des environnements à forte volumétrie et charges de travail intensives

Comment les plateformes SaaS et API assurent une montée en charge fiable dans le cloud



Évolutivité des charges de travail à forte intensité de données : approche concrète

Ce cas d'usage propose une approche concrète et technique de l'évolutivité pour ce type de charges, avec notamment :

1

Les principaux défis rencontrés à mesure que les systèmes se développent

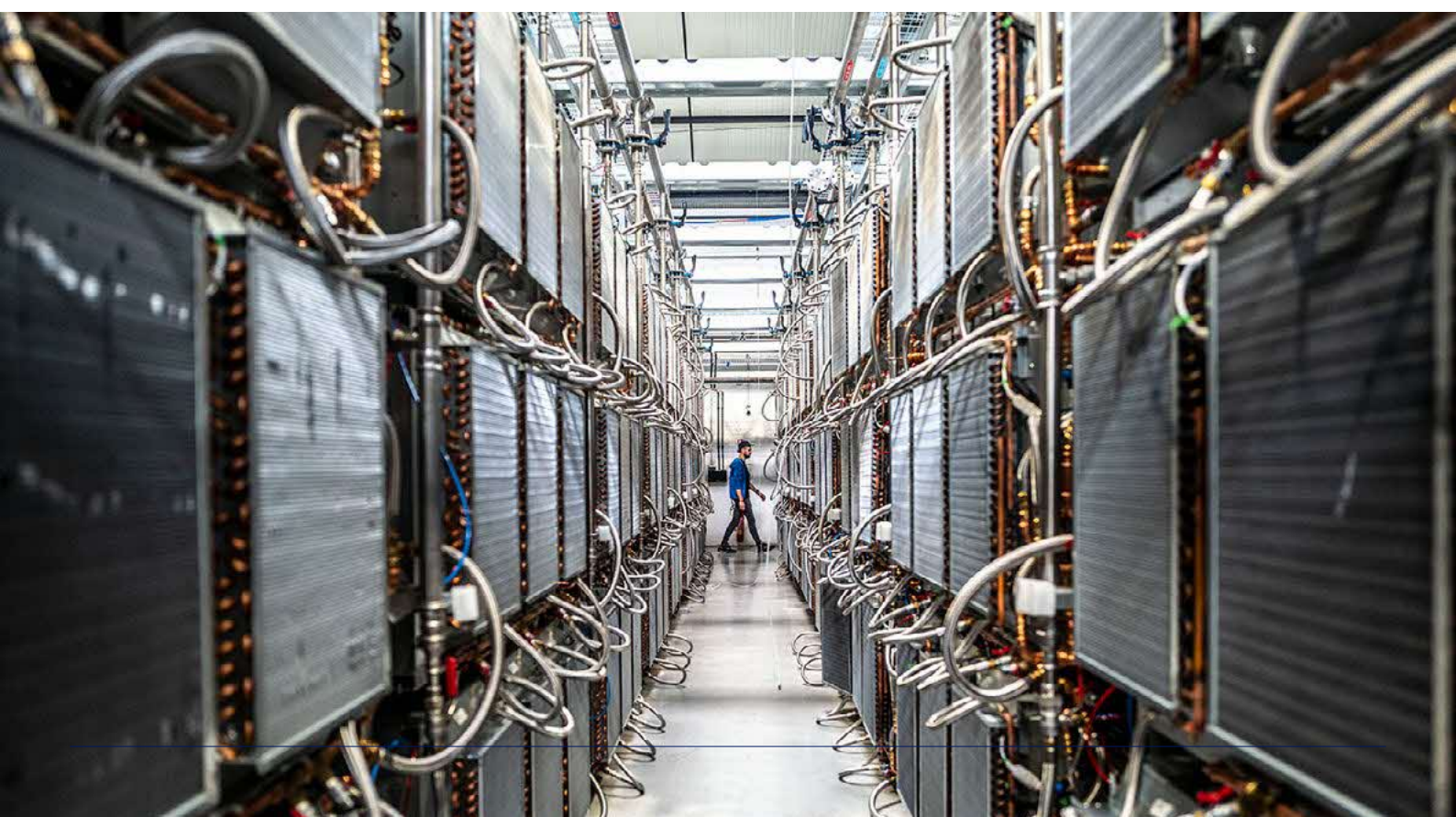
2

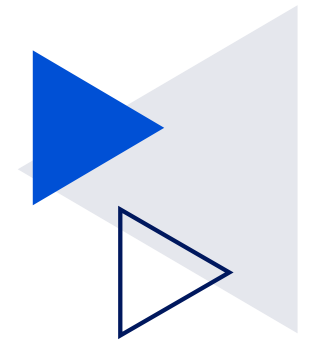
Une liste de recommandations techniques pour orienter les choix d'infrastructure

3

Des exemples de déploiement illustrant comment les organisations assurent performance, efficacité et maîtrise

Ces éléments permettront aux ingénieurs et architectes de mieux comprendre les schémas et stratégies favorisant une montée en charge fiable, tout en maintenant un bon niveau de contrôle sur la complexité et les coûts.





1. Défis techniques de l'évolutivité des systèmes à forte volumétrie de données

À mesure que les volumes de données, le débit et la concurrence augmentent, de nouveaux défis techniques apparaissent.



Limites de performance du stockage, notamment lorsque les ensembles de données passent de quelques gigaoctets à plusieurs téraoctets ou plus



Difficulté à anticiper les coûts de croissance du stockage, des entrées/sorties (I/O) et des exports de données



Maintien d'une faible latence, y compris en période de forte charge, pour les opérations d'ingestion, les requêtes et les appels API



Complexité opérationnelle accrue liée à la gestion de bases de données distribuées et des pipelines



Manque de flexibilité pour faire évoluer indépendamment le stockage, les bases de données et la capacité de calcul, entraînant une inefficacité opérationnelle



Contraintes de localisation des données et exigences de conformité, augmentant la complexité architecturale



Goulots d'étranglement sur le réseau liés à un trafic est-ouest important entre services

Pour répondre efficacement à ces enjeux, il est essentiel de s'appuyer sur une infrastructure offrant une visibilité et un contrôle accrus sur le stockage, le transfert, le traitement et la mise à l'échelle des données. Cela permet de garantir des performances stables et des coûts prévisibles, en adéquation avec l'évolution de la plateforme.

2. Liste de recommandations techniques pour l'évolutivité

Avant de valider une architecture ou de sélectionner une plateforme cloud, il est important d'évaluer le comportement réel de la charge de travail en environnement de production. Des ensembles de données importants, des opérations intensives en I/O et un volume de requêtes élevé exercent une

pression spécifique sur le stockage, le réseau et les systèmes de traitement distribué.

La liste ci-dessous identifie les caractéristiques techniques clés à prendre en compte pour choisir une solution cloud ou une architecture adaptée aux besoins de votre organisation.



Stockage

- ▶ Type de stockage nécessaire : objet, bloc, fichier ou combinaison
- ▶ IOPS maximum requis
- ▶ Besoin éventuel de disques NVMe pour les charges à faible latence
- ▶ Priorité de performance : débit soutenu ou accès aléatoire en lecture/écriture
- ▶ Prévisions de croissance des données (Go/To/Po)
- ▶ Modèle de réplication et de durabilité : mono-zone, multi-zone de disponibilité ou multi-région



Calcul et traitement

- ▶ Profil de la charge de travail : contraintes liées au processeur, à la mémoire ou au recours au traitement GPU
- ▶ Mode de traitement des données : traitement par lot, en flux continu ou en temps réel
- ▶ Exigences liées au traitement distribué : Clusters Spark, Dask, ClickHouse ou autres solutions similaires
- ▶ Modèle d'usage : charge stable, à pics, ou déclenchée par événements
- ▶ Capacité de conteneurisation et d'auto-scaling



Réseau

- ▶ Sensibilité à la latence
- ▶ Besoins en bande passante est-ouest entre services ou nœuds
- ▶ Nécessité d'une architecture multi-zone ou multi-région
- ▶ Besoin d'un réseau privé
- ▶ Volumes de données sortantes (egress) prévus



Évolutivité

- ▶ Modèle de montée en charge : horizontale (stateless), verticale (augmentation de débit), ou hybride
- ▶ Fréquence et typologie des pics de trafic : quotidiens, saisonniers ou liés à des événements spécifiques
- ▶ Exigences en matière d'auto-scaling, de sharding ou d'extension de cluster
- ▶ Degré de maturité vis-à-vis de la conteneurisation



Coûts et gestion des ressources

- ▶ Prévisibilité de la consommation : usage stable ou très variable
- ▶ Principaux postes de coûts : stockage, puissance de calcul, trafic sortant ou montée en charge des bases de données
- ▶ Modèle de facturation souhaité : forfait mensuel ou paiement à l'usage
- ▶ Impact des volumes d'objets stockés et du modèle de réplication sur les coûts globaux



Conformité et sécurité

- ▶ Exigences en matière de localisation et de souveraineté des données
- ▶ Contraintes réglementaires relatives à la localité des données
- ▶ Certifications attendues
- ▶ Besoins en matière de chiffrement, d'isolement réseau et d'intégration avec les systèmes IAM
- ▶ Attentes clients ou réglementaires concernant la confidentialité et la conformité



Automatisation et déploiement

- ▶ Nécessité d'outils d'Infrastructure-as-Code
- ▶ Intégration aux chaînes CI/CD existantes
- ▶ Délai de mise à disposition des environnements
- ▶ Accès via API et déclencheurs (hooks) d'automatisation
- ▶ Exigences en matière de supervision, journalisation, alertes et observabilité



Considérations relatives aux bases de données

- ▶ Type de base de données : documentaire (MongoDB) ou relationnelle (PostgreSQL)
- ▶ Contraintes de lecture/écriture en parallèle
- ▶ Modèle de requêtes attendu : transactionnel ou analytique
- ▶ Exigences en matière de haute disponibilité, de supervision, de sauvegarde et de mécanismes de basculement



Des bénéfices concrets et mesurables pour les équipes au quotidien

Des plateformes correctement dimensionnées, conçues en tenant compte du comportement réel des charges de travail actuelles et à venir, offrent aux équipes d'ingénierie des gains significatifs en matière de fiabilité, de performance et

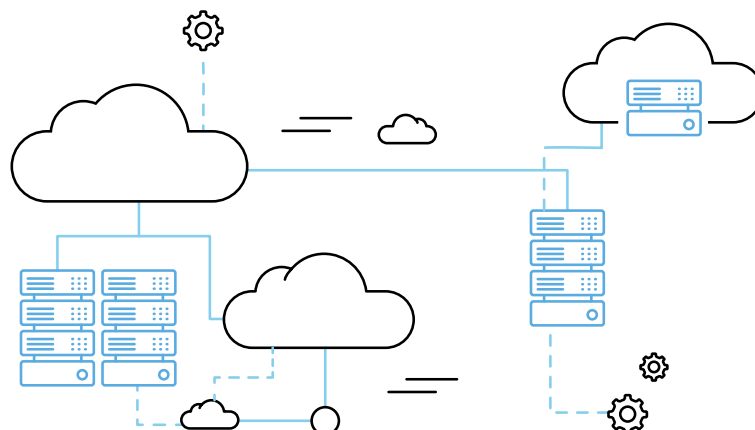
d'efficacité. Elles permettent notamment de garantir un niveau de service constant, d'adapter les ressources de manière indépendante et de mieux maîtriser les services ainsi que les coûts associés.

CAPACITÉ

- ✓ Débit d'ingestion stable
- ✓ Réduction des temps de traitement et de requête
- ✓ Évolutivité indépendante du stockage
- ✓ Complexité opérationnelle réduite
- ✓ Coûts liés au cloud maîtrisés
- ✓ Maîtrise des données et de la conformité
- ✓ Architecture ouverte

GAINS OPÉRATIONNELS

- ✓ Moins de pertes de données, engagements de niveau de service (SLA) respectés
- ✓ Latence API optimisée, réponses analytiques accélérées
- ✓ Possibilité de faire évoluer séparément le stockage, la puissance de calcul et les bases de données
- ✓ Réduction du temps dédié aux mises à jour, à la haute disponibilité et aux sauvegardes
- ✓ Moins de frais inattendus liés aux IOPS ou à l'égress
- ✓ Meilleur contrôle sur la localisation et le traitement des données
- ✓ Moins de dépendance fournisseur, meilleure portabilité



Architecture pour les charges de travail IA

L'essor de l'IA accélère l'émergence de charges de travail intensives en données, qui combinent des besoins élevés en calcul, un débit de stockage soutenu et une exigence forte en matière d'efficacité. Face à cette évolution, les équipes doivent s'appuyer sur une infrastructure garantissant des performances prévisibles, un déploiement flexible et une cohérence opérationnelle, sans ajout de complexité inutile.

Une approche pragmatique consiste à normaliser un socle technologique compatible avec des environnements hybrides.

Des plateformes comme OVHcloud incarnent cette approche en proposant une combinaison d'infrastructures prêtes à déployer sur site (On-Prem Cloud Platform), de serveurs bare metal et d'instances cloud traditionnelles, souvent construites sur des technologies haute performance comme AMD EPYC.

Les équipes peuvent ainsi sélectionner l'infrastructure la plus adaptée à chaque type de charge, tout en bénéficiant d'un environnement familier, d'un comportement maîtrisé et d'une efficacité à grande échelle.

3. Des déploiements SaaS et API conçus pour l'évolutivité

Les plateformes SaaS et API à forte volumétrie de données font face à différents défis d'infrastructure en fonction du trafic, de la taille des ensembles de données et du type de charge de travail.

Les cas suivants illustrent comment des entreprises en croissance parviennent à maintenir des performances constantes, à assurer une montée en charge efficace et à garder le contrôle sur leurs dépenses.

Si votre défi est...

Conformité réglementaire et haute disponibilité multi-AZ

Traitement et stockage de données à grande échelle

Voir...

iATROS
MapTiler

iATROS : une plateforme de santé numérique sécurisée, conforme et hautement disponible

iATROS exploite une plateforme de santé numérique destinée à collecter, analyser et restituer des données sensibles de patients, utilisées par plusieurs centaines de milliers d'utilisateurs. Pour répondre à des exigences réglementaires strictes (incluant le RGPD et des normes sectorielles), tout en optimisant la

latence pour des utilisateurs répartis à travers l'Europe, l'entreprise a revu l'architecture de sa plateforme autour d'une infrastructure cloud multi-clusters, avec haute disponibilité et sécurité renforcée.

DÉFI PRINCIPAL

- ▶ Migrer depuis une infrastructure non conforme aux exigences européennes en matière de sécurité des données
- ▶ Garantir une faible latence et une disponibilité continue pour des utilisateurs répartis sur plusieurs zones géographiques
- ▶ Respecter les exigences de conformité RGPD, ISO et spécifiques au secteur de la santé
- ▶ Assurer une résilience sur plusieurs zones de défaillance

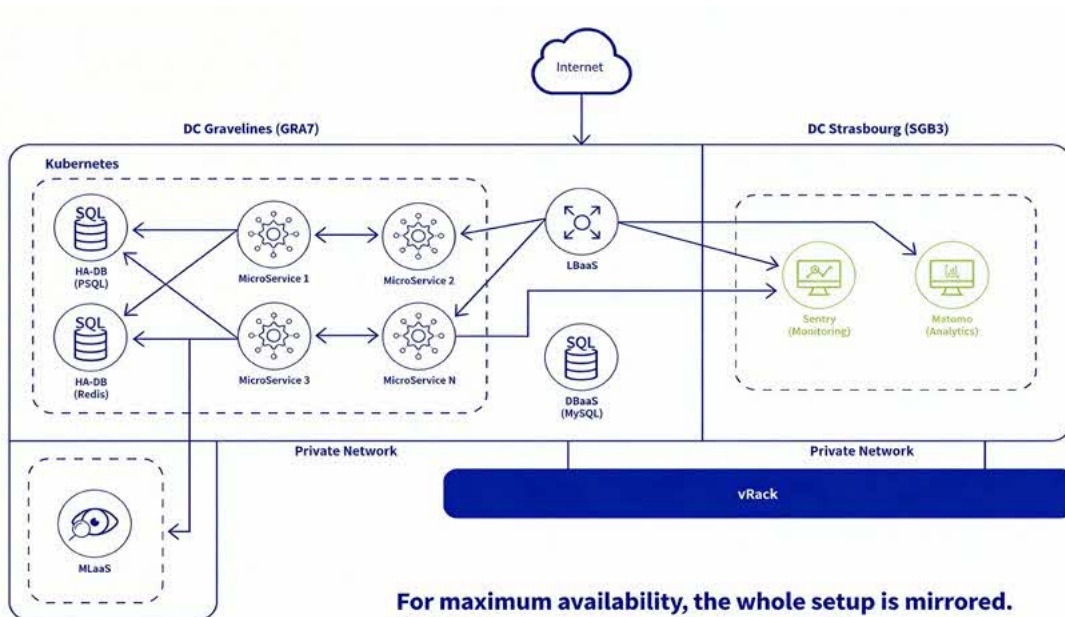
SOLUTION MISE EN PLACE

- ▶ Migration vers des datacenters OVHcloud situés en UE pour garantir la conformité RGPD
- ▶ Déploiement de clusters PostgreSQL en haute disponibilité répartis sur plusieurs zones (multi-AZ)
- ▶ Intégration du réseau privé vRack pour optimiser le trafic est-ouest
- ▶ Utilisation de bases de données managées et d'instances évolutives pour garantir des performances constantes
- ▶ predictable performance

RÉSULTATS

- ▶ Réduction d'environ 20 % des ressources utilisées et des coûts par rapport à l'ancienne solution
- ▶ Amélioration sensible de la latence pour l'ensemble des utilisateurs
- ▶ Hébergement sécurisé et pleinement conforme au RGPD, avec une gouvernance robuste et une architecture résiliente

[Lire l'étude de cas complète](#)



MapTiler : production de cartes satellites évolutives grâce à un nombre illimité d'instances cloud

MapTiler, scale-up suisse spécialisée dans les technologies géospatiales, développe et distribue des fonds de cartes haute performance ainsi que des données cartographiques personnalisées. Ses solutions sont utilisées dans des domaines variés tels que la logistique, l'immobilier, la défense ou encore

le tourisme. Pour rester compétitive en fournissant des images satellites actualisées et en répondant à plusieurs centaines de millions de vues cartographiques mensuelles, MapTiler recherchait une infrastructure cloud capable de monter en charge sans créer de points de blocage.



DÉFI PRINCIPAL

- ▶ Gérer une croissance rapide du stockage liée à l'accumulation continue de données satellites
- ▶ Répondre de manière fiable à environ 400 millions de requêtes quotidiennes de tuiles cartographiques
- ▶ Éviter les coûts variables et difficiles à anticiper associés à des charges de données fluctuantes
- ▶ Réduire significativement les délais de traitement, passant de plusieurs années à quelques semaines



SOLUTION MISE EN PLACE

- ▶ Migration vers le Public Cloud OVHcloud avec un stockage d'objets compatible S3 hautement évolutif
- ▶ Utilisation d'un nombre illimité d'instances cloud pour le traitement parallèle des données
- ▶ Adoption d'un modèle de tarification prévisible sur le stockage et le calcul
- ▶ Rationalisation de la production cartographique via une infrastructure optimisée

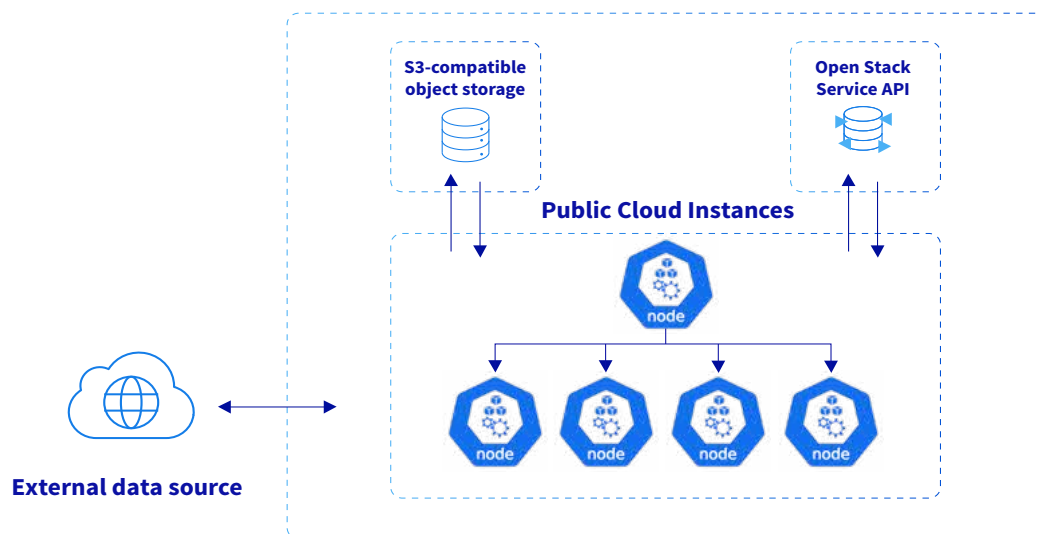


RÉSULTATS

- ▶ Diminution du temps de traitement de données satellites estimé de plus de 18 ans à seulement quelques semaines
- ▶ Coûts mensuels stabilisés, quel que soit le volume de données traité
- ▶ Élimination des limites de capacité grâce à l'allocation dynamique d'instances de calcul

[Lire l'étude de cas complète](#)

L'infrastructure cloud de MapTiler chez OVHcloud



*S3 is a registered trademark of Amazon Technologies, Inc. OVHcloud services are not sponsored or approved by, nor affiliated with Amazon Technologies, Inc. in any way.

Faire évoluer les charges de travail intensives en données tout en gardant le contrôle

À mesure que les charges de travail à forte intensité de données se développent, les enjeux opérationnels s'intensifient avec la montée en charge des volumes, des services et des utilisateurs. Sans une conception appropriée, des goulots d'étranglement de performance, une hausse des coûts et une surcharge opérationnelle peuvent apparaître. Qui plus est, souvent au moment critique où la plateforme est le plus sollicitée.

Les organisations les plus résilientes veillent à permettre une montée en charge indépendante des ressources clés : stockage, calcul, réseau et bases de données. Combinée à l'usage de services managés pour le stockage, l'orchestration ou la supervision, cette approche permet aux équipes de se concentrer sur les charges métier essentielles tout en

maintenant une flexibilité opérationnelle.

Une infrastructure conçue pour l'évolutivité à long terme permet de garder la maîtrise, même lorsque :

- ▶ les volumes de données augmentent de façon exponentielle ;
- ▶ le nombre de services ou de pipelines se multiplie ;
- ▶ la fréquence des incidents techniques devient un enjeu.
- ▶

Une infrastructure cloud flexible permet de limiter les frictions et d'accompagner la croissance, sans compromis sur la performance.

Faites évoluer vos plateformes SaaS et API à fort volume de données en toute confiance.

Découvrez la différence OVHcloud.

Envie d'en savoir plus ?

Planifiez un échange avec un architecte solutions :

[Demander un appel](#)

Explorez un cloud conçu pour accompagner la croissance des entreprises :

[En savoir plus](#)