



SUMMIT  
EN LIGNE

# Mieux exploiter Amazon EC2

Jonathan Dion

Évangéliste technique principal  
Amazon Web Services

# Programme

Les principes de base d'Amazon EC2

Calcul pour tous les type de charges de travail

Optimisation des coûts

Optimisation de capacité

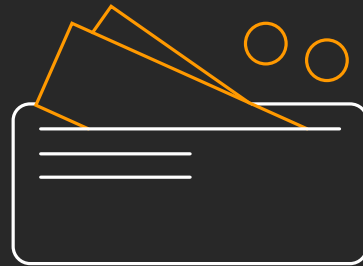
Recommandations

Exemples de charges de travail

# Amazon EC2 il y a plus de 13 ans



Taille unique



Payez pour ce que  
vous utilisez



Augmentez ou  
diminuez la capacité  
rapidement

# La plateforme de choix la plus large et la plus importante

## Charges de travail

Usage général  
Capacité extensible  
Forte intensité de calcul  
Forte intensité de mémoire  
Stockage (E/S élevées)  
Stockage dense  
Calcul GPU  
Forte intensité graphique  
Inférence



## Capacités

Choix du processeur  
(AWS, Intel, AMD)  
Processeur rapide  
(jusqu'à 4.0 GHz)  
Taille de mémoire élevée  
(jusqu'à 24 TiB)  
Stockage d'instances  
(HDD et NVMe)  
Calcul accéléré  
(GPUs et FPGA)  
Réseau  
(jusqu'à 100 Gbps)  
Bare Metal  
Taille  
(Nano à 32xlarge)



## Options

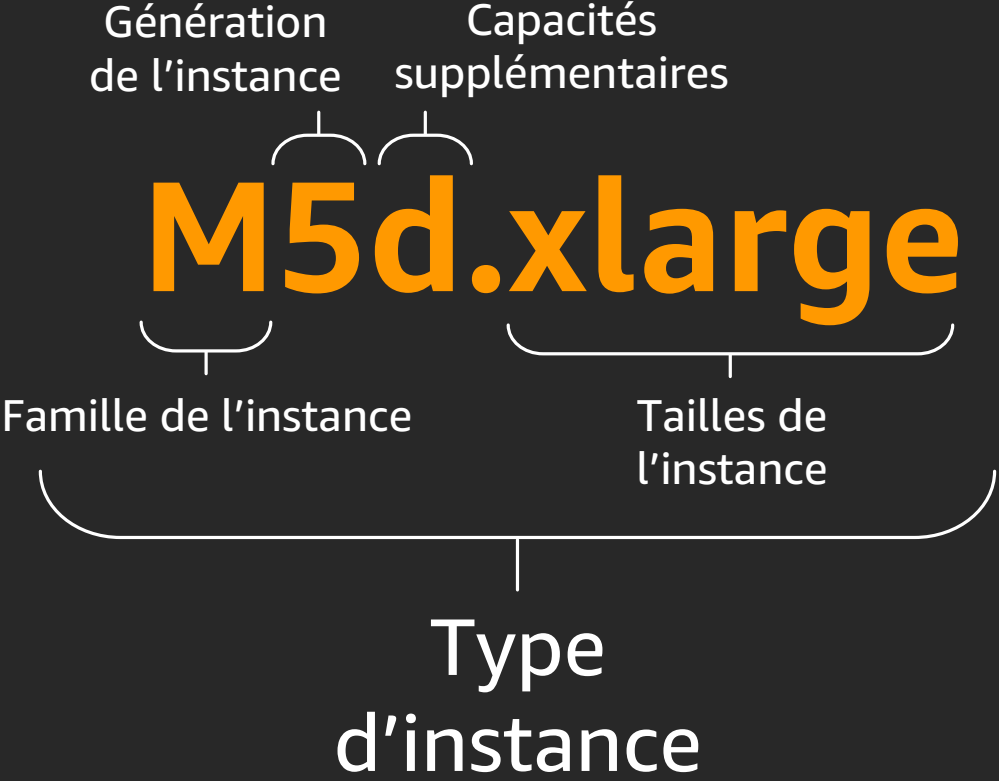
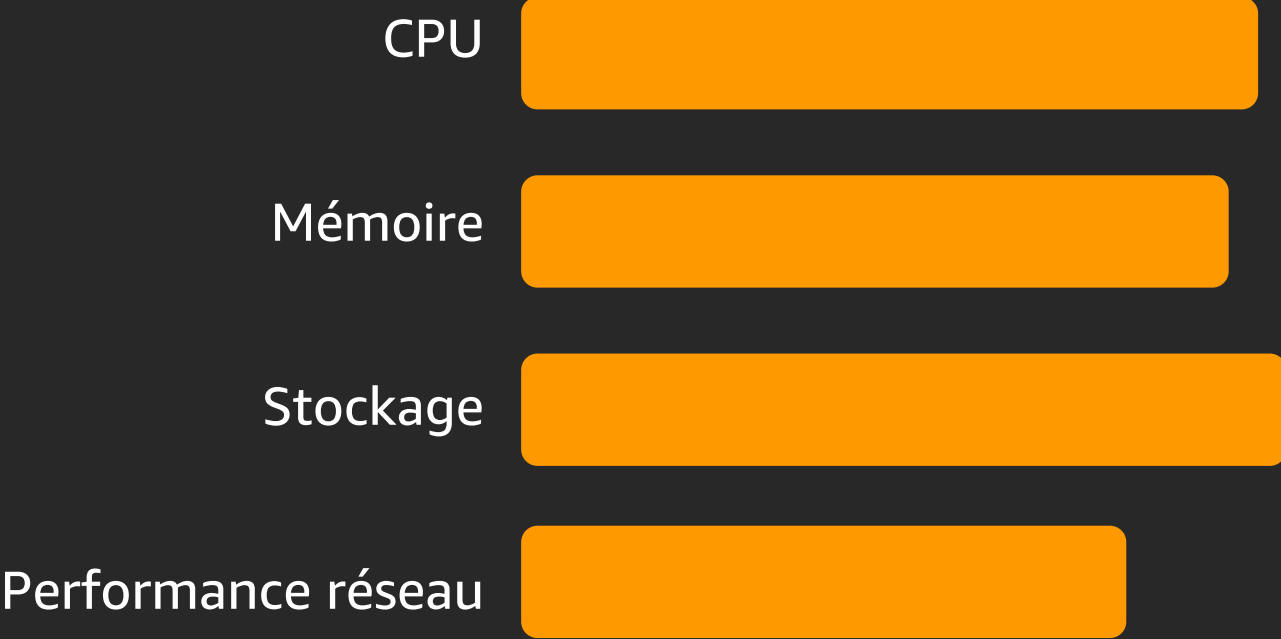
Amazon Elastic Block Store  
(Amazon EBS)  
Amazon Elastic Graphics  
Amazon Elastic Inference



# Plus de 275

types d'instances  
pour pratiquement  
toutes les charges  
de travail et tous  
les besoins  
professionnels

# Caractéristiques des instances Amazon EC2



# Les instances à usage général d'Amazon EC2

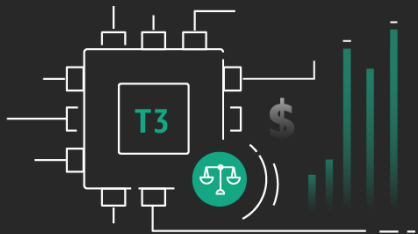


## Instances M5

Un ensemble équilibré de ressources de calcul, de mémoire et de réseau

Un rapport mémoire à vCPU de 4:1

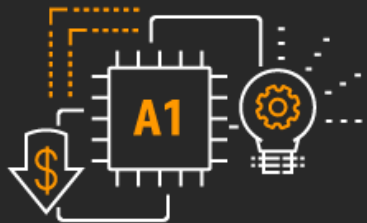
---



## Instances T3

Fournit un niveau de base de performance du CPU avec la capacité d'aller au-dessus de la base pour les charges de travail qui ne nécessitent pas une performance en continue

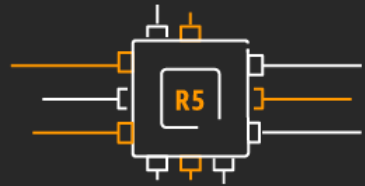
---



## Instances A1

Charges de travail pouvant être réparties sur plusieurs cœurs, pouvant tenir dans la mémoire disponible, et fonctionner sur Arm

# Instances de mémoire optimisée Amazon EC2



## Instances R5

Accélérez les performances des charges de travail qui traitent de grands ensembles de données en mémoire

Un rapport mémoire à vCPU de 8:1

---



## Instances X1 et X1e

Pour les charges de travail exigeantes en mémoire et les charges de travail en mémoire très importantes

Un rapport mémoire à vCPU entre 16:1 et 32:1

---



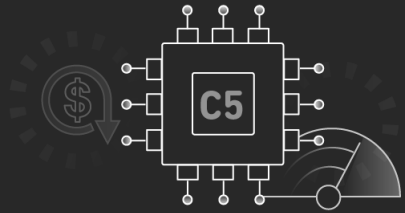
## Instances Mémoire élevée

Besoins en mémoire

Certifiées pour le fonctionnement de SAP HANA

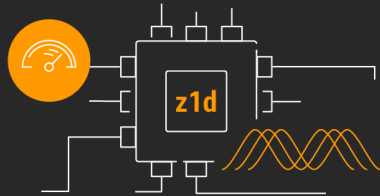
De 6 To à 24 To de mémoire

# Instances de calcul optimisé Amazon EC2



## Instances C5

Des performances élevées à faible coût pour un rapport mémoire à vCPU de 2:1



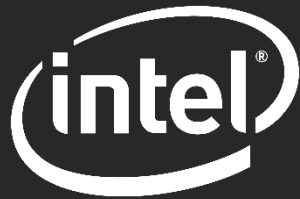
## Instances z1d

Des performances élevées par thread

Processeur le plus rapide du cloud avec une vitesse de 4.0 GHz

Un rapport mémoire à vCPU de 8:1

# Le plus grand choix de type processeurs



---

Processeur  
Intel Xeon  
Scalable



---

Processeur  
AMD EPYC



---

Processeur  
Graviton

# Annonce des processeurs AWS Graviton2

## Processeur Graviton1



Le premier processeur basé sur Arm dans tous les cloud majeurs

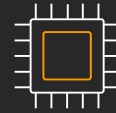


Dotés de cœurs Arm Neoverse 64 bits avec 16-nm semi-conducteur conçu par AWS



Jusqu'à 16 vCPU, un réseau de 10 Gbps, 3,5 Gbps de bande passante Amazon EBS

## Processeur Graviton2



Dotés de cœurs Arm Neoverse de 64 bits; 7-nm semi-conducteur conçu par AWS



Jusqu'à 64 vCPU, un réseau de 20 Gbps, 14 Gbps de bande passante Amazon EBS



7x plus performant, 4x plus de vCPU et 5x plus de mémoire

# 6 nouvelles instances utilisant AWS Graviton2

## Usage général

4 GB DRAM/vCPU

**M6g**

**M6gd**

## Calcul optimisé

2 GB DRAM/vCPU

**C6g**

**C6gd**

## Mémoire optimisée

8 GB DRAM/vCPU

**R6g**

**R6gd**

---

Pour un aperçu, contactez-nous

À venir en 2020

---

À venir en 2020

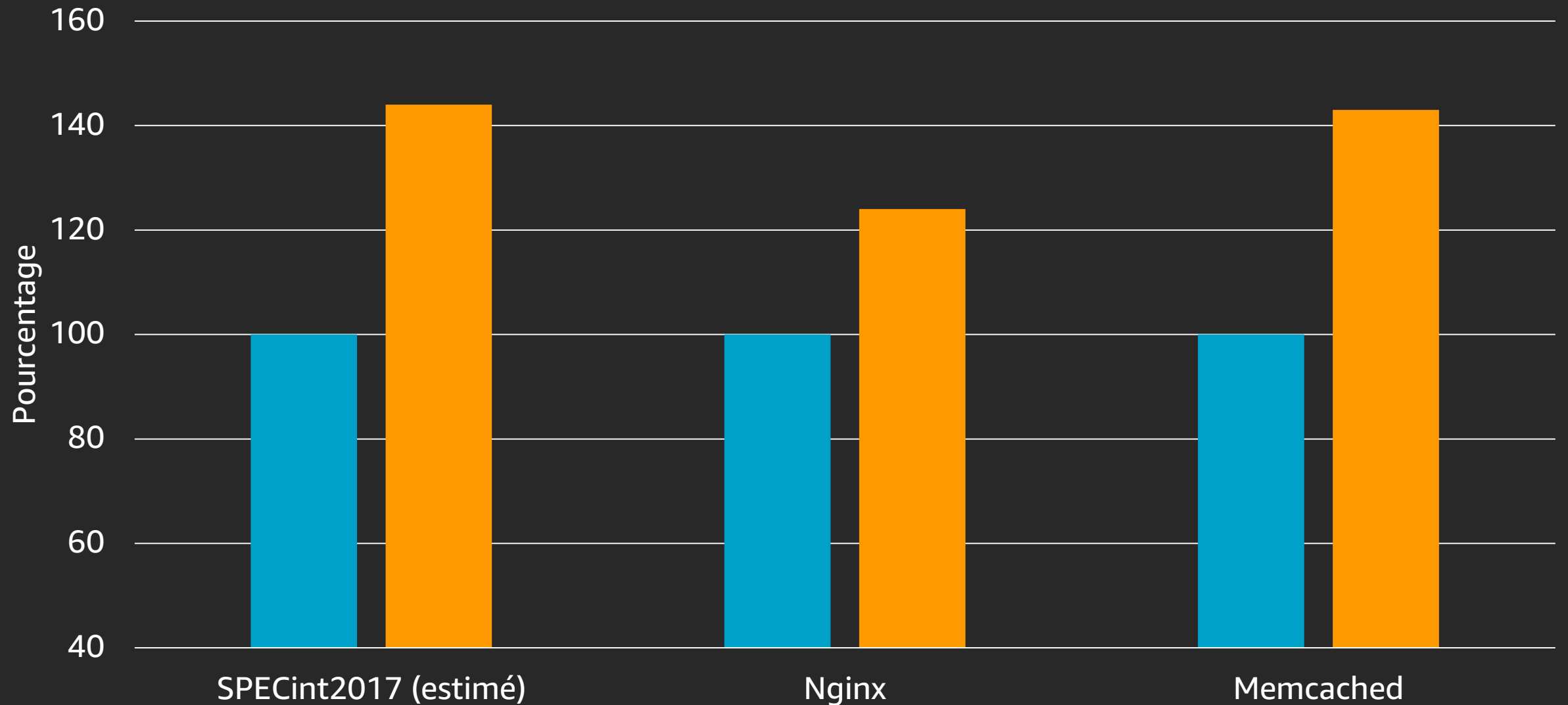
Tous avec une mise en réseau améliorée, Amazon EBS et 3 avec des SSD NVMe

# Standards de l'industrie et charges de travail

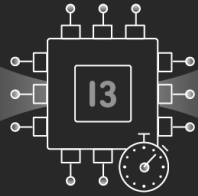
5<sup>e</sup> génération x86



M6g



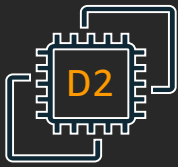
# Instances d'Amazon EC2 avec stockage optimisé



Instances  
I3/I3en

E/S optimisées pour les charges de travail élevées et celles à faible latence

---



Instances  
D2

Coût le plus bas par stockage (\$/GB)  
Prise en charge d'un débit séquentiel élevé

---



Instances  
H1

Conçues pour les applications nécessitant un faible coût, une capacité de débit de disque élevée et des opérations E/S séquentielles rapides pour les ensembles de données importants

Davantage de vCPU et de mémoire par To de stockage que l'instance D2

# Instances à calcul accéléré Amazon EC2

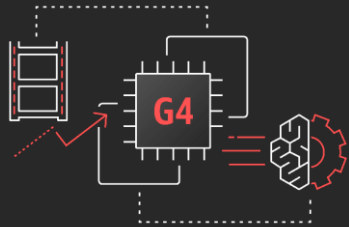


## Série P

Instances P2/P3

Instances de **calcul** GPU conçues pour l'apprentissage profond, les simulations en calcul de haute performance, l'informatique financière, et les rendus par lots.

Intègre des GPU NVIDIA haut de gamme, dont le Volta V100

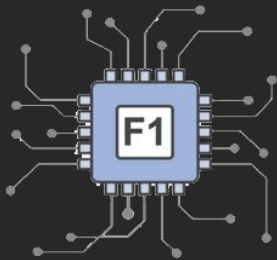


## Série G

Instances G3/G4

Instances **graphiques** GPU conçues pour le rendu 3D, les postes de travail virtuels, l'encodage vidéo et l'AR/VR

Intègre des GPU NVIDIA de milieu de gamme, tels que les GPU Turing T4, avec les fonctionnalités et la licence GRID pour les postes de travail virtuels



## Instances FPGA

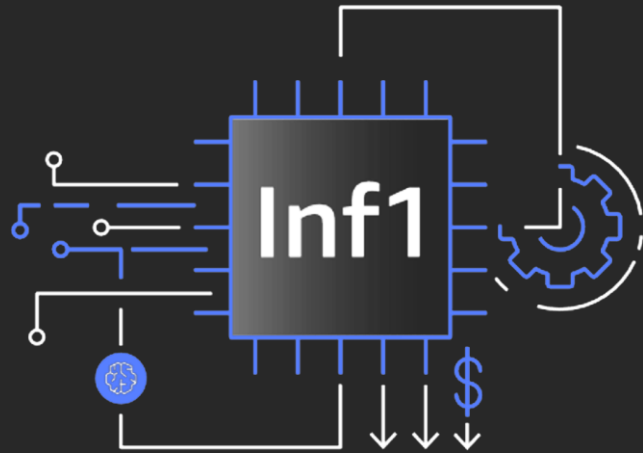
Instances F1

**FPGA programmables** par les clients offrant des améliorations de performances impressionnantes pour des applications telles que l'informatique financière, la génomique, la recherche accélérée et le traitement d'image

FPGA Xilinx Virtex UltraScale+ VU9P en une seule instance  
Programmable par VHDL, Verilog ou OpenCL

# Annonce des instances Inf1

## Annonce des instances Inf1



**Inférence d'apprentissage  
automatique de haute  
performance au coût le plus bas  
dans le cloud**

Des coûts par inférence jusqu'à 40 % inférieurs à ceux des instances Amazon EC2 GPU

Débit d'inférence jusqu'à 3x plus élevé avec jusqu'à 2 000 TOPS à une latence inférieure à la milliseconde

Intégration avec les cadres d'apprentissage automatique les plus populaires, notamment TensorFlow, PyTorch, et MXNet

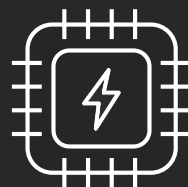
# Commence par nos investissements dans la plateforme AWS Nitro System

## Carte Nitro



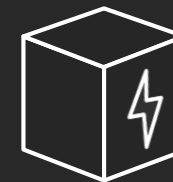
Stockage local NVMe  
Amazon EBS  
Réseau, surveillance et sécurité

## Puce de sécurité Nitro



Intégrée à la carte mère  
Protège les ressources matérielles

## Hyperviseur Nitro

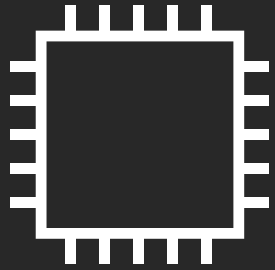


Hyperviseur léger  
Allocation de mémoire et CPU  
Performances similaires à celles de *Bare Metal*

---

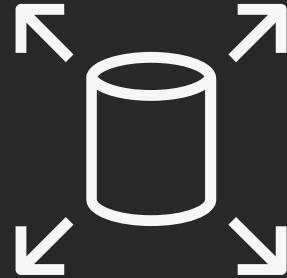
Blocs de construction modulaires pour une conception et une livraison rapide des instances EC2

# Gamme de stockage en mode bloc



## Stockage d'instances

Stockage temporaire de niveau bloc attaché à l'hôte, idéal pour le stockage d'informations qui changent fréquemment ou sont répliquées sur plusieurs instances



## Amazon EBS

Simple d'utilisation, conçu pour l'utilisation avec Amazon Elastic Compute Cloud (EC2) pour les charges de travail exigeantes en termes de débit et d'opérations



## Instantanés

Sauvegardes incrémentielles et ponctuelles de vos données Amazon EBS pouvant être utilisées pour restaurer de nouveaux volumes, étendre la taille d'un volume ou déplacer des volumes dans d'autres zones de disponibilité

# Nouvelles améliorations des performances et de la sécurité EBS

Chiffrement par défaut pour les volumes EBS avec paramètre d'activation



Chiffrer tous les nouveaux volumes EBS dans une région avec un seul paramètre

Facilité à garantir la conformité sans modification des charges de travail

Fast snapshot restore (FSR)

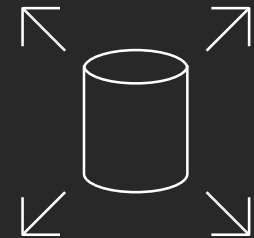


Objectif de temps de récupération (RTO) 6 fois plus faible

Passer l'étape du préchauffage - accès instantané aux données et pleine puissance lors de la création du volume

Restaurez jusqu'à 10 volumes simultanément

Bande passante EBS optimisée 36 % plus élevée sur les types d'instances C5/C5d, M5/M5d et R5/R5d



Bande passante dédiée à Amazon EBS

Bande passante maximale de 19 Gbps, la plus élevée parmi les instances EC2

# Optimiser les coûts et la capacité d'Amazon EC2

```
elif operation == "MIRROR_Y":
    mirror_mod.use_x = False
    mirror_mod.use_y = True
    mirror_mod.use_z = False
elif operation == "MIRROR_Z":
    mirror_mod.use_x = False
    mirror_mod.use_y = False
    mirror_mod.use_z = True

#selection at the end -add back the deselected mirror modifier object
mirror_ob.select= 1
modifier_ob.select=1
bpy.context.scene.objects.active = modifier_ob
print("Selected" + str(modifier_ob)) # modifier ob is the active ob
#mirror_ob.select = 0
done = bpy.context.selected_objects[0]
bpy.data.objects[me.name].select = 1
```

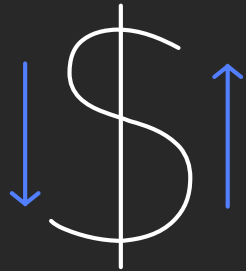
AJK5545001J-JK

AD-58457-DJ-JK

# Optimiser les coûts et la capacité d'Amazon EC2

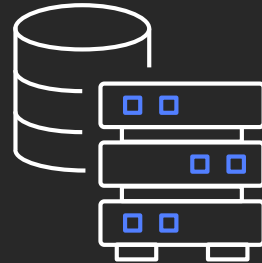
Nous continuons d'innover pour nos clients

## Coût



Obtenir un prix et des performances optimales avec différents modèles d'achat

## Capacité



La gestion de la capacité est facile sur la plateforme de calcul la plus large et la plus importante

## Recommandations

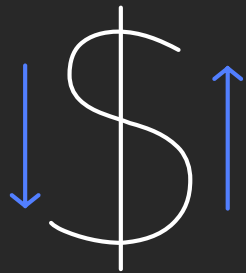


Les recommandations en matière de coût et de capacité permettent de faciliter l'utilisation et de gagner du temps

# Optimiser les coûts et la capacité d'Amazon EC2

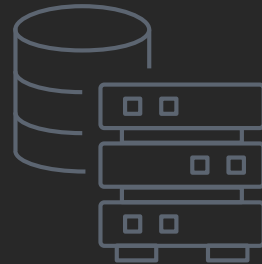
Nous continuons d'innover pour nos clients

## Coût



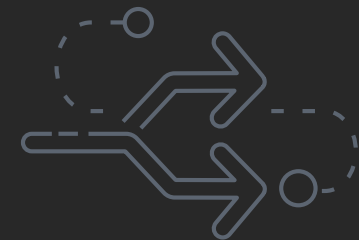
Obtenir un prix et des performances optimales avec différents modèles d'achat

## Capacité



La gestion de la capacité est facile sur la plateforme de calcul la plus large et la plus importante

## Recommandations

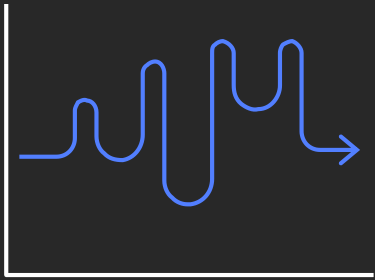


Les recommandations en matière de coût et de capacité permettent de faciliter l'utilisation et de gagner du temps

# Options d'achat pour Amazon EC2

## À la demande

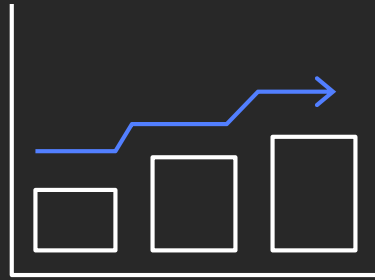
Paiement de capacité de calcul **par seconde** et sans engagement long terme



Charges de travail irrégulières pour définir les besoins

## Instances réservées (RIs)

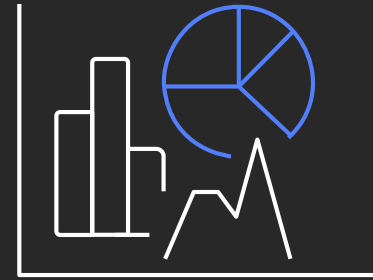
Avec un engagement de 1 ou 3 ans, vous recevez une **remise importante** par rapport au tarif des instances à la demande



Une utilisation engagée et d'état stable

## Savings Plans

Mêmes remises qu'avec les instances réservées (RIs) avec **plus de flexibilité**



Un accès flexible au calcul

## Instances Spot

Capacités réservées d'Amazon EC2 **pour économiser jusqu'à 90 %** du prix des instances à la demande



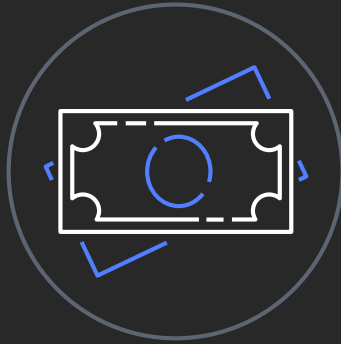
Des charges de travail tolérantes aux pannes, flexibles et statiques

# Introduction des Savings Plans



## Facile d'utilisation

Recevez automatiquement des taux réduits en échange d'un engagement monétaire



## Remises importantes

Choisissez parmi deux types de Savings Plans pour bénéficier de réductions allant jusqu'à 72 % pour les EC2 Instance Savings Plans et 66 % pour les Compute Savings Plans



## Flexibles

Prenez un engagement qui s'applique à plusieurs services de calcul AWS, même si vos besoins changent

L'option d'achat flexible offre jusqu'à 72 % de réduction sur l'utilisation d'Amazon EC2, de AWS Fargate et de AWS Lambda

# Types de Savings Plans



## Compute Savings Plans

Offre la plus grande flexibilité. Jusqu'à 66 % de réduction (mêmes prix que les instances réservées convertibles)

### Flexible à travers de

- ✓ Famille d'instance : ex. Passer de C5 à M5
- ✓ Région : par ex. changement de région depuis l'US Ouest (Oregon) vers Canada (Centre)
- ✓ OS : ex. Windows vers Linux
- ✓ Location : ex. passage de la location dédiée à la location par défaut
- ✓ Options de calcul : ex. passage de EC2 vers Fargate



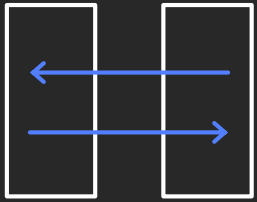
## EC2 Instance Savings Plans

Fournit les prix les plus bas. Jusqu'à 72 % de réduction (comme pour les instances réservées standards) sur la famille d'instances sélectionnée (par exemple, C5 ou M5), dans une région AWS spécifique

### Flexible à travers de

- ✓ Taille : ex. passer de m5.xl à m5.4xl
- ✓ OS : ex. passer de m5.xl sous Windows à m5.xl sous Linux
- ✓ Location : ex. passer de la location dédiée m5.xl à la location par défaut m5.xl

# Économisez jusqu'à 90 % en utilisant les instances EC2 Spot



## Instances

Même infrastructure que les instances à la demande et réservées



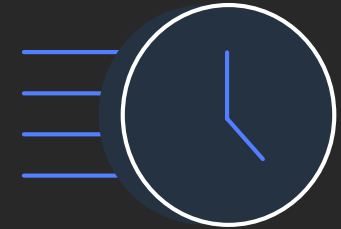
## Coût

Changements fluides et peu fréquents ; plus prévisible



## Utilisation

Choisissez différents types et tailles d'instances et AZ dans une seule flotte ou un groupe AutoScaling EC2

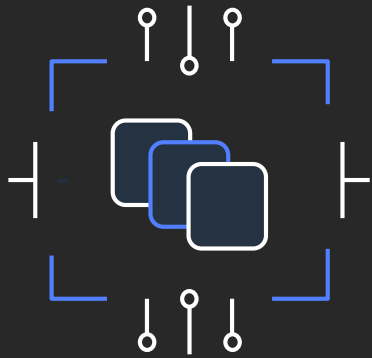


## Capacité

Les interruptions ne se produisent que si le service à la demande a besoin de capacité

**Les prix sont basés sur les tendances à long terme de l'offre et de la demande ; pas besoin de faire un enchère !**

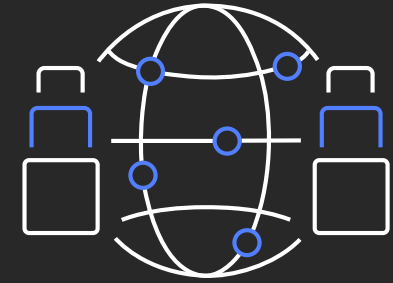
# La flexibilité est la clé d'une utilisation réussie de Spot



Flexibilité en termes  
d'instances



Flexibilité de temps



Flexibilité au niveau des  
régions

# Gérer les interruptions de Spot

Moins de 5 % des instances Spot ont rencontré des interruptions durant les 3 derniers mois.

## Interruptions minimales



Surveillez l'avis d'interruption de 2 minutes via les métadonnées d'instance ou les événements Amazon CloudWatch, et automatisez en

- ✓ faisant des points de contrôle
- ✓ drainant depuis ELB
- ✓ utilisant la fonction démarrer-arrêter ou hiberner afin de redémarrer plus rapidement

---

## Gestionnaires d'interruption pour Amazon ECS et Amazon EKS



Amazon  
EKS



Amazon  
ECS

- ✓ Connexion entre les demandes d'interruption de l'infrastructure AWS et les nodes
- ✓ Les tâches exécutées sur les instances Spot seront automatiquement arrêtées avant l'arrêt de l'instance, et les tâches de remplacement seront programmées ailleurs dans le cluster

# Optimiser les coûts et la capacité d'Amazon EC2

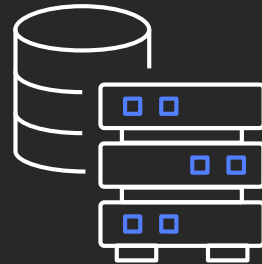
Nous continuons d'innover pour nos clients

## Coût



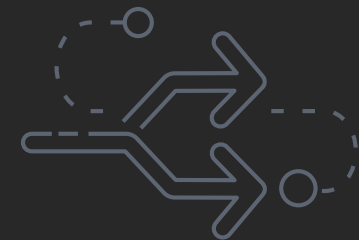
Obtenir un prix et des performances optimales avec différents modèles d'achat

## Capacité



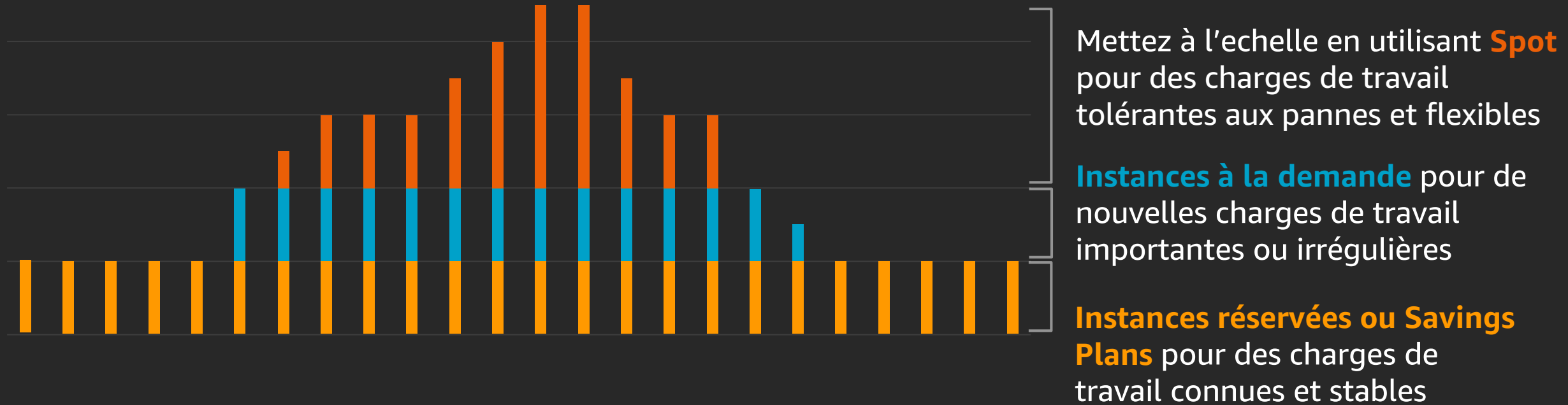
La gestion de la capacité est facile sur la plateforme de calcul la plus large et la plus importante

## Recommandations



Les recommandations en matière de coût et de capacité permettent de faciliter l'utilisation et de gagner du temps

# Pour optimiser Amazon EC2, combinez les options d'achat



# Utiliser Amazon EC2 Auto Scaling

Automatiquement mettre les instances à l'échelle à travers des familles d'instances et des options d'achat dans un seul groupe de Auto Scaling group afin d'optimiser les coûts



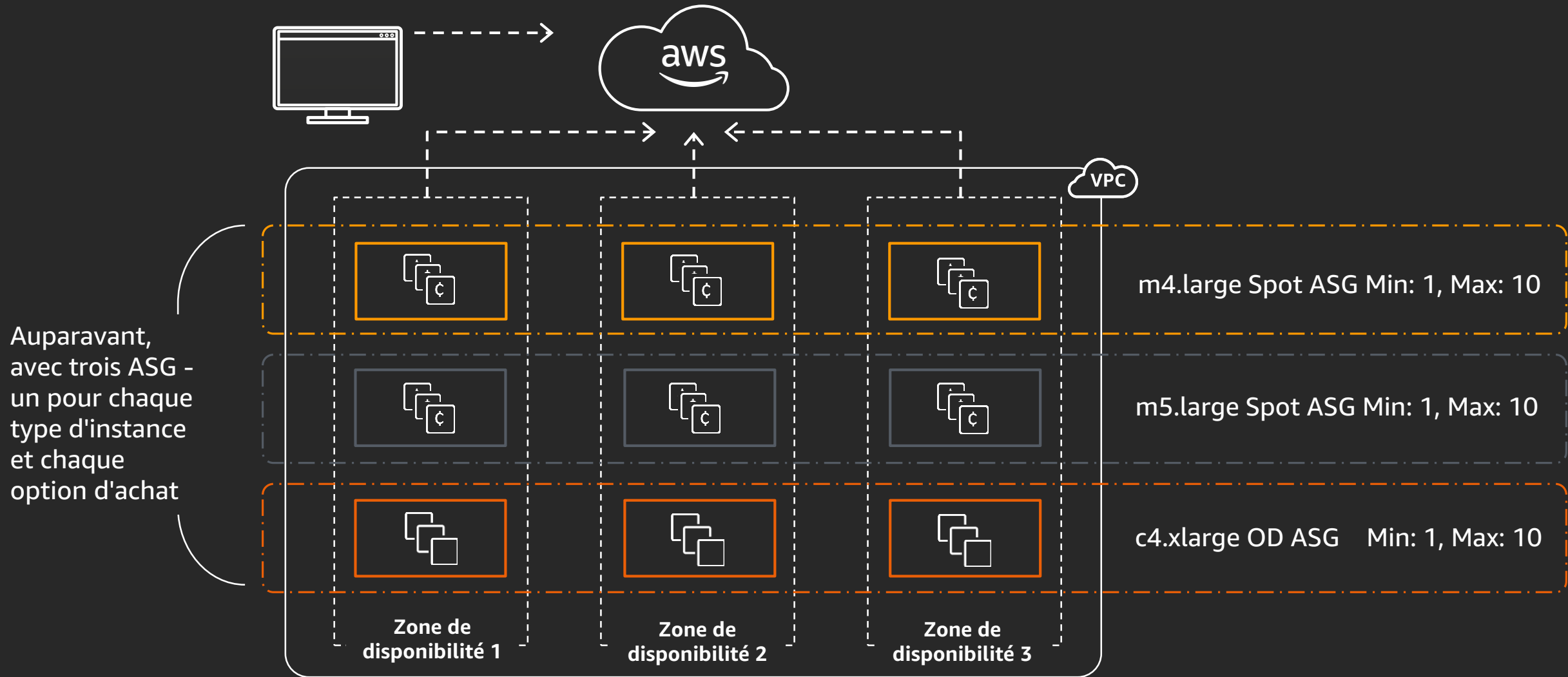
Atelier : <https://ec2spotworkshops.com/running-amazon-ec2-workloads-at-scale.html>

Réduire le coût

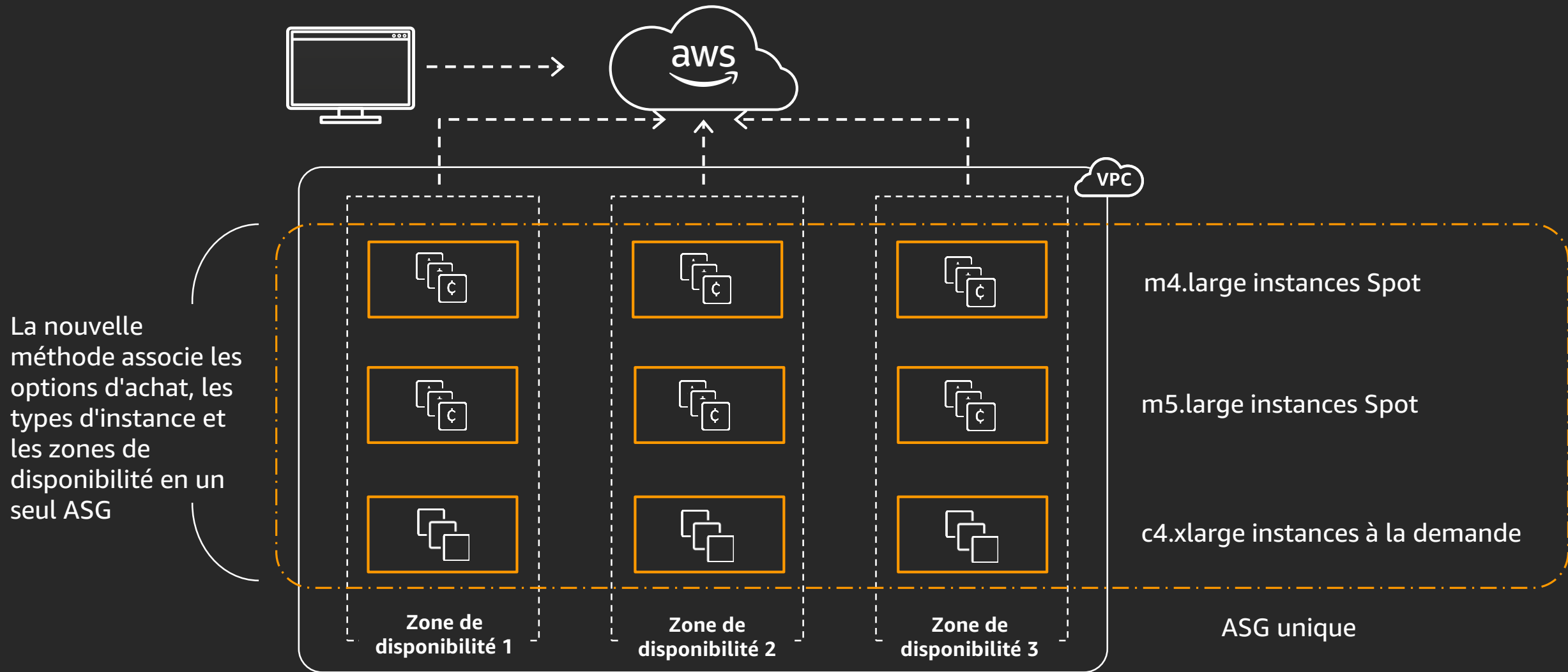
Optimiser la performance

Éliminer les surcoûts opérationnels

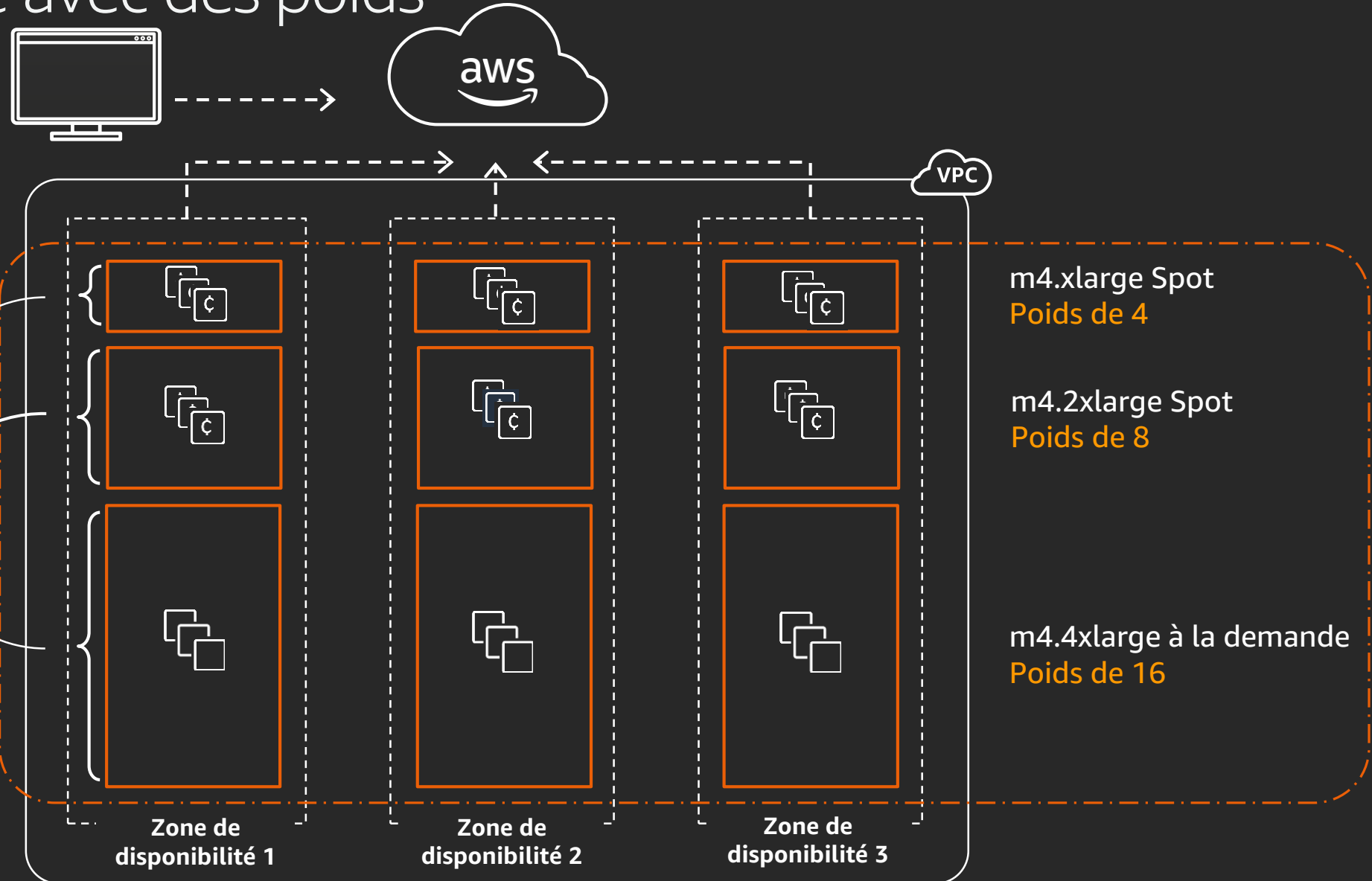
# Auparavant : plusieurs ASG pour utiliser Spot ou à la demande et les instances réservées ensemble



# Ensuite : Spot, à la demande et les instances réservées dans un ASG unique



# Maintenant : Spot, à la demande et les instances réservées dans un ASG unique avec des poids



# Utiliser Amazon EC2 Auto Scaling

Automatiquement mettre les instances à l'échelle à travers des familles d'instances et des options d'achat dans un seul groupe de Auto Scaling group afin d'optimiser les coûts

## Capacité optimisée

Privilégier le déploiement des instances Spot dans un pool Spot de plus grande taille afin de réduire les risques d'interruption

## Le coût le plus bas

Prioriser les coûts en sélectionnant une combinaison d'instances à la demande et Spot lancées en se basant sur le prix le plus bas

## Liste de priorités

Utiliser une liste de priorités pour les types d'instances à la demande afin d'augmenter la capacité lors d'un événement urgent et imprévisible



Atelier : <https://ec2spotworkshops.com/running-amazon-ec2-workloads-at-scale.html>

Réduire le coût

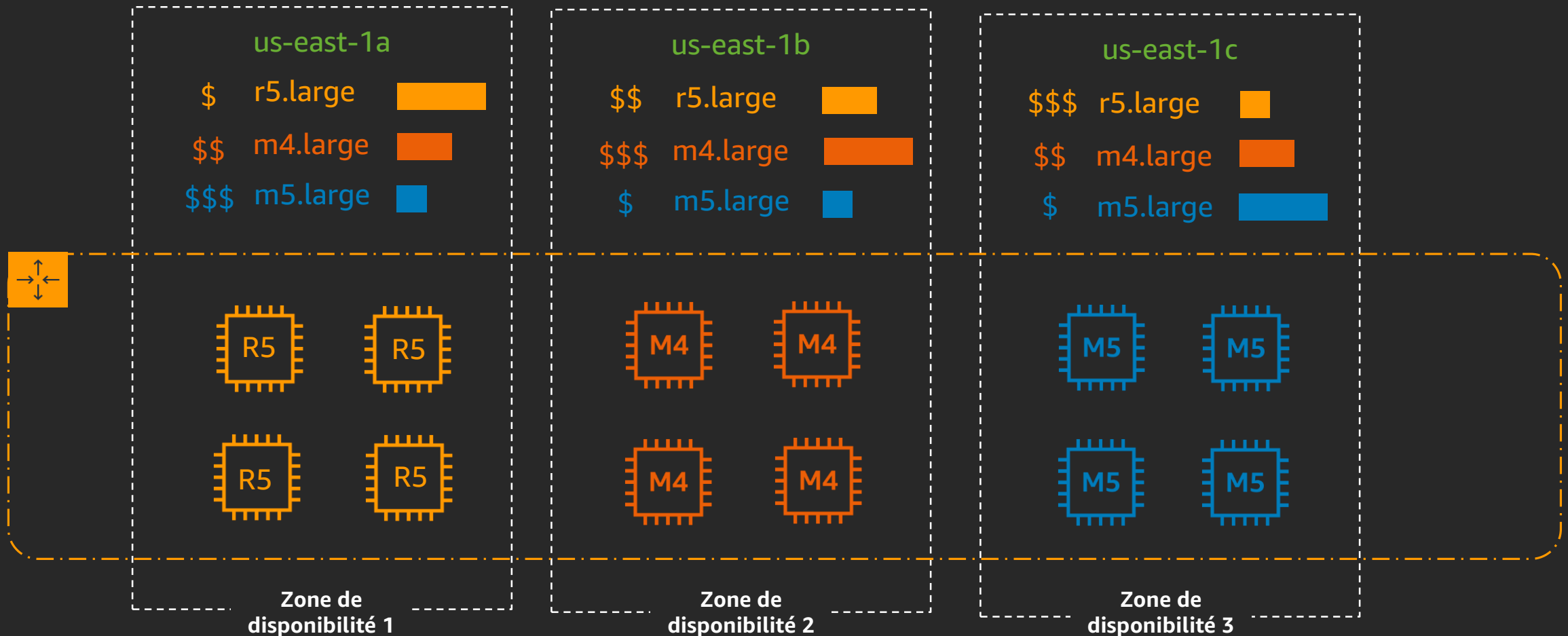
Optimiser la performance

Éliminer les surcoûts opérationnels

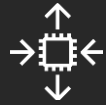
# Stratégie d'attribution de capacités optimisées de l'ASG

Desired capacity: 12    OnDemandBaseCapacity: 0    OnDemandPercentageAboveCapacity: 0

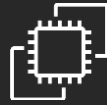
overrides: ["r5.large", "m4.large", "m5.large"]    SpotAllocationStrategy: **capacity-optimized**



# Intégration avec les instances Spot et EC2 Auto Scaling



Amazon EC2  
Auto Scaling



Flotte  
Amazon EC2



AWS  
Thinkbox



Amazon  
EMR



AWS  
CloudFormation



AWS  
Batch



Amazon  
ECS



Amazon  
EKS



Amazon  
SageMaker



AWS  
Fargate



AWS Elastic  
Beanstalk



# Optimiser les coûts et la capacité d'Amazon EC2

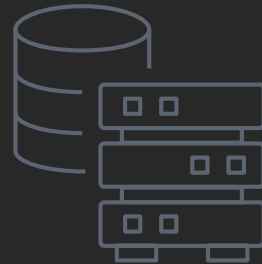
Nous continuons d'innover pour nos clients

## Coût



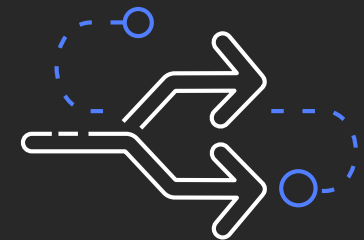
Obtenir un prix et des performances optimales avec différents modèles d'achat

## Capacité



La gestion de la capacité est facile sur la plateforme de calcul la plus large et la plus importante

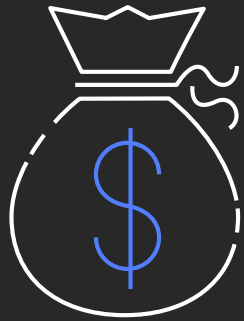
## Recommandations



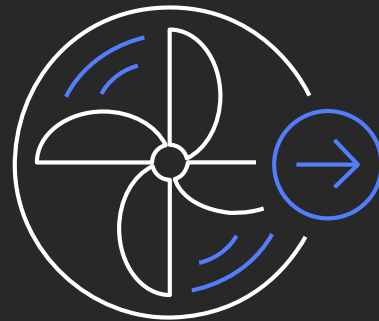
Les recommandations en matière de coût et de capacité permettent de faciliter l'utilisation et de gagner du temps

# AWS Compute Optimizer

Recommande les instances optimales pour Amazon EC2 et les groupes Amazon EC2 Auto Scaling sur plus de 140 instances des familles M, C, R, T, et X



Réduit les **coûts** et améliore la **performance** de la charge de travail



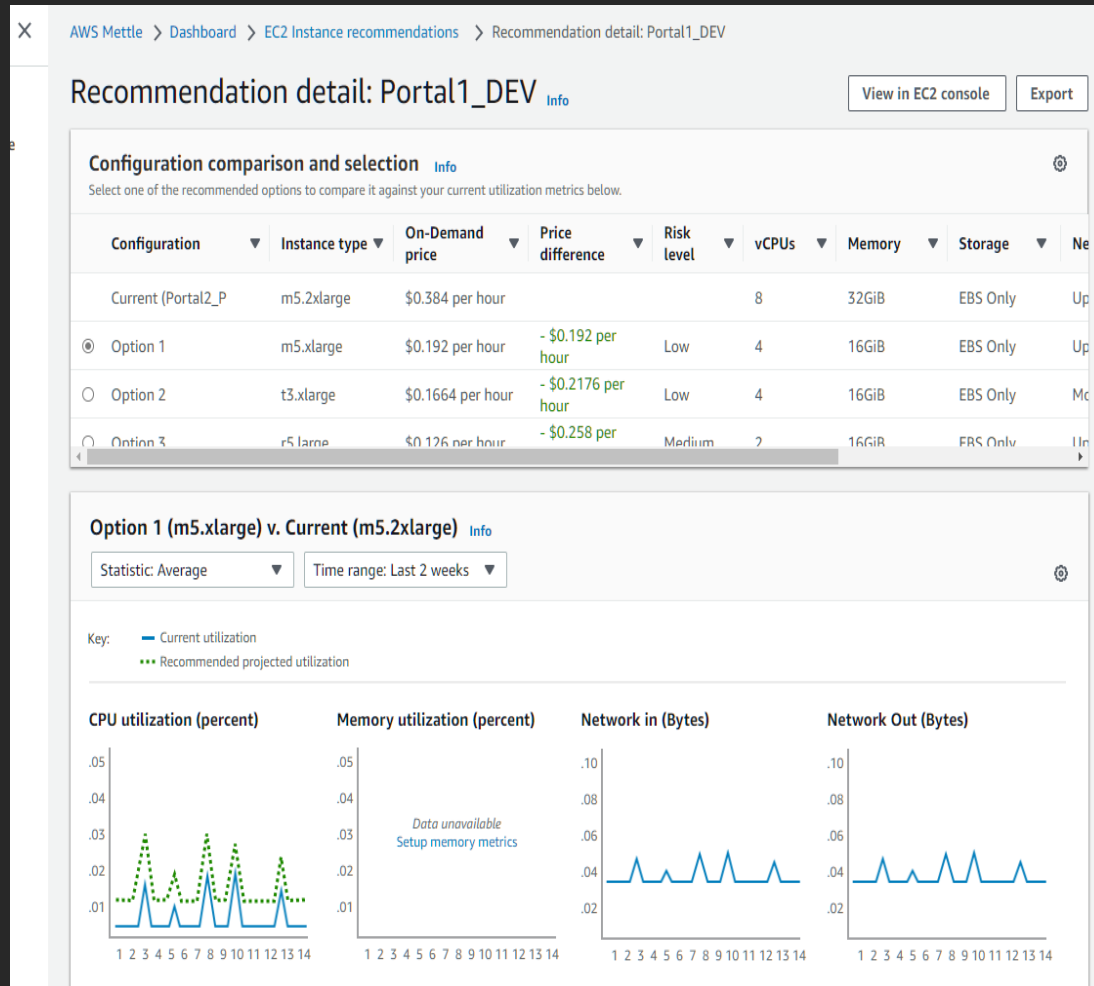
**Applique les enseignements** tirés de millions de charges de travail pour formuler des recommandations



**Permet de gagner du temps** en comparant et en sélectionnant les ressources optimales pour votre charge de travail

# Facile de faire un choix avec AWS Compute Optimizer

De nouveaux services qui recommandent des ressources de calcul AWS optimales pour une réduction des coûts jusqu'à 25 %.



Recommande les instances EC2 optimales

Optimise les performances et réduit les coûts en formulant des recommandations pour vous aider à adapter vos calculs à votre charge de travail

Analyse les métriques d'Amazon CloudWatch et prend en compte la configuration de groupe Auto Scaling pour des recommandations intuitives et réalisables

Jusqu'à trois recommandations par charge de travail

Disponible sans frais supplémentaires

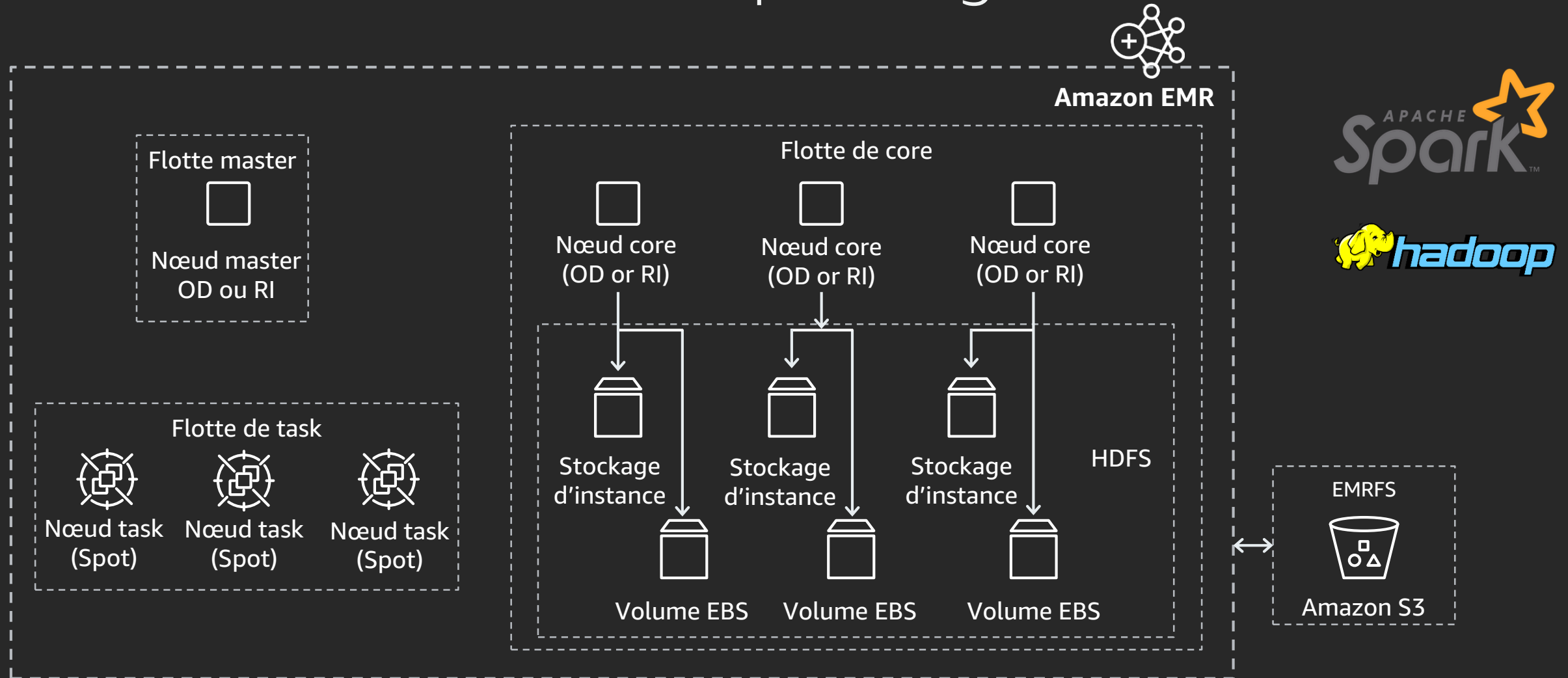
Les charges  
de travail  
sur AWS

**Analytique et *Big Data***

DevOps - (CI/CD)

Applications Web

# Architecture de référence pour Big data



Atelier : [https://ec2spotworkshops.com/running\\_spark\\_apps\\_with\\_emr\\_on\\_spot\\_instances.html](https://ec2spotworkshops.com/running_spark_apps_with_emr_on_spot_instances.html)

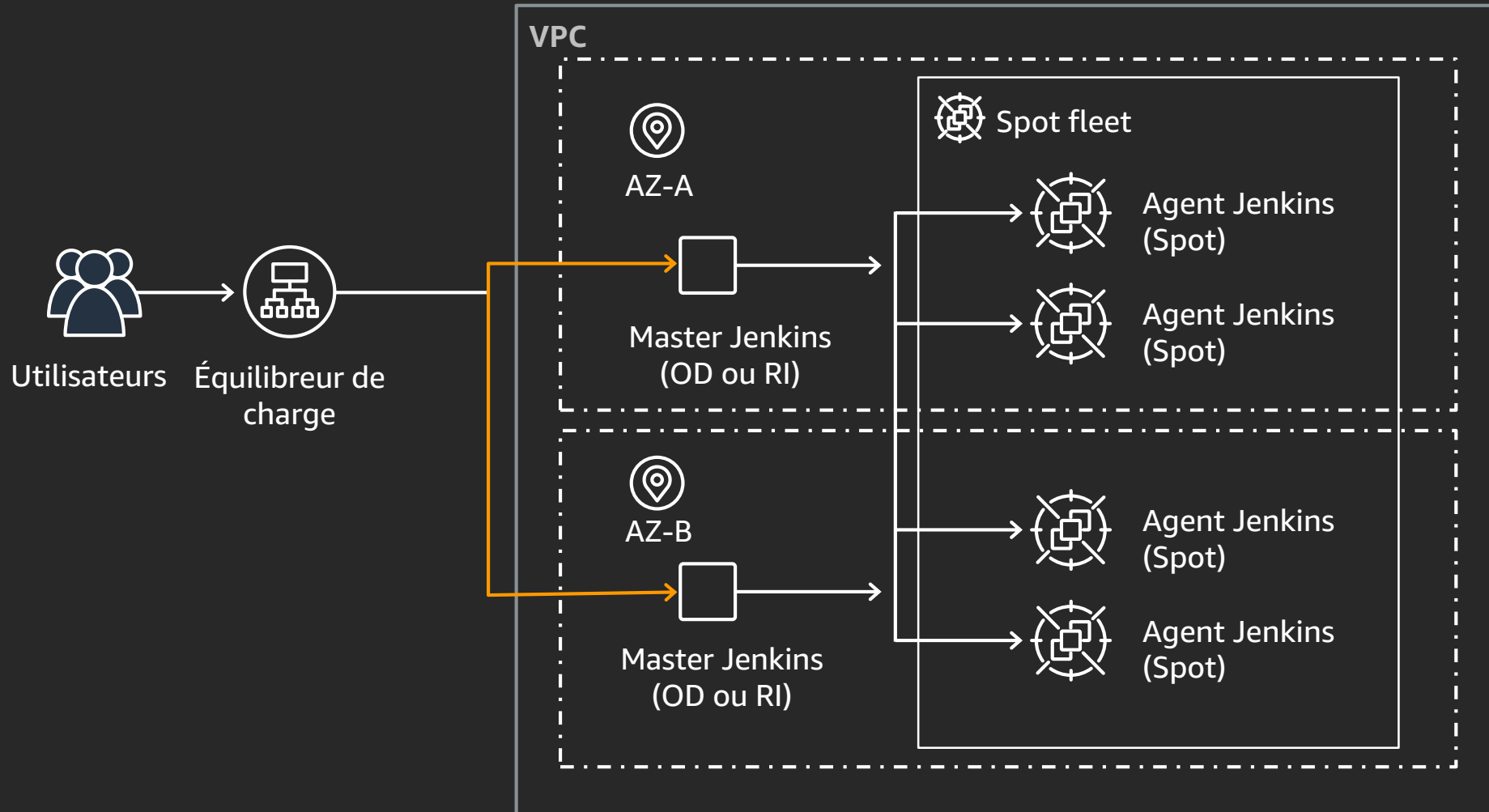
# Les charges de travail sur AWS

Analytique et Big Data

**DevOps - (CI/CD)**

Applications Web

# Architecture de référence CI/CD



<https://github.com/aws-labs/ec2-spot-jenkins-plugin/>

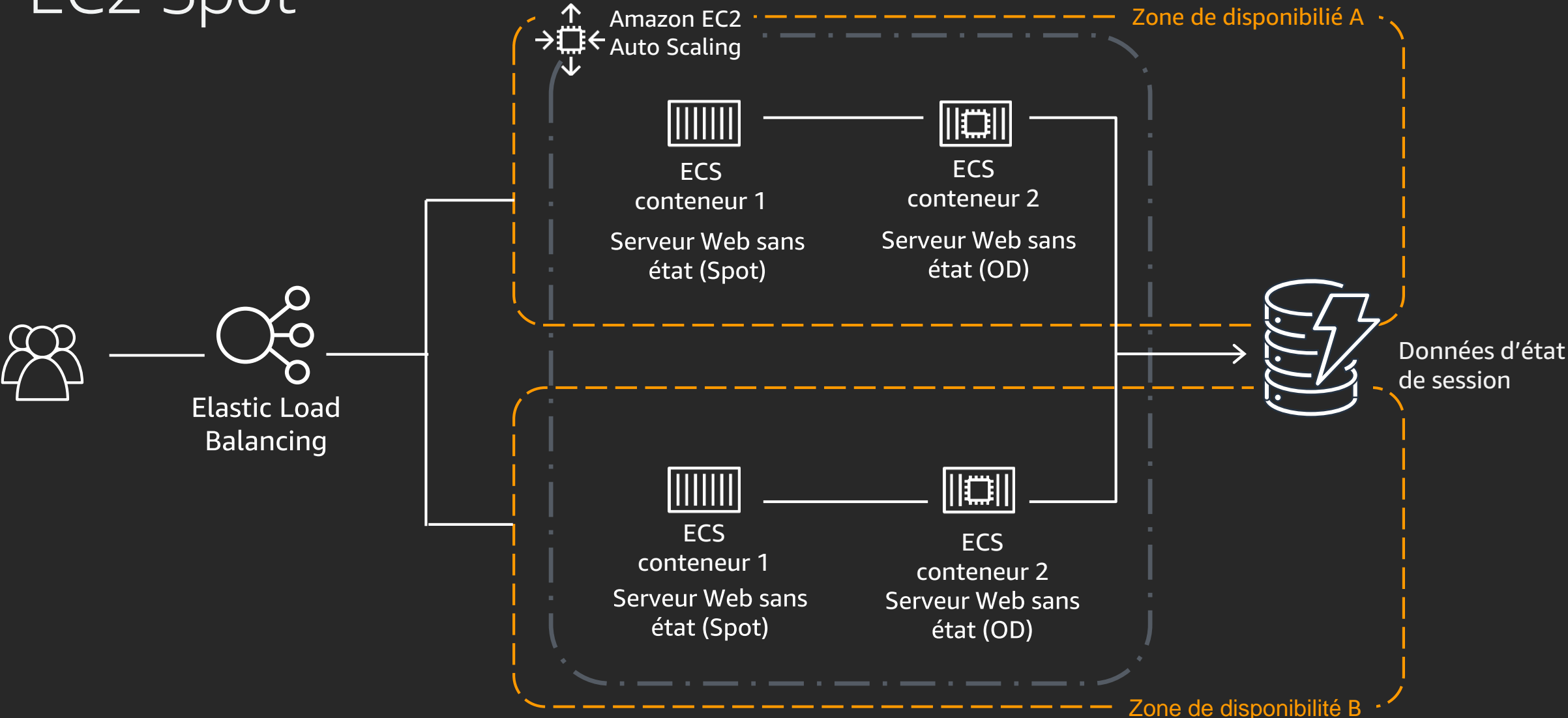
# Les charges de travail sur AWS

Analytique et Big Data

DevOps - (CI/CD)

**Applications Web**

# Exécuter des applications Web avec Amazon ECS sur EC2 Spot



Pour conclure...



# Les points clés à retenir

1

EC2 a le bon type d'instance pour chaque charge de travail

**Charge de travail optimisée pour des instances EC2, AWS Nitro System, Elastic Block Storage**

2

Accéder à du calcul à un coût inférieur pour innover rapidement

**Les instances Spot et Savings Plans**

3

Comment automatiser l'optimisation des coûts et la capacité

**EC2 Auto Scaling**

4

Optimisez votre charge de travail grâce aux bonnes pratiques

**AWS Compute Optimizer**

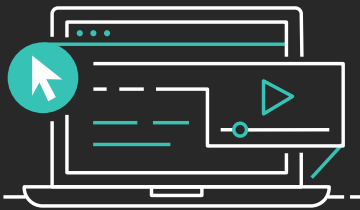
5

Obtenez des conseils techniques lors d'une journée en immersion AWS

**CI/CD, analytique, Big Data et les services Web**

# Apprenez le calcul avec la formation et la certification AWS

Des ressources créées par les experts d'AWS pour vous aider à développer vos compétences en matière de cloud computing



Plus de 20 cours en ligne gratuits parcourent des sujets liés au cloud computing, y compris des introductions aux services suivants :

- Amazon EC2
- Amazon EC2 Auto Scaling
- AWS Systems Manager
- Amazon EC2 Inf1 instances



Le calcul est aussi couvert dans le cours offert en classe et en français **Architecture sur AWS**. Vous aurez accès à des instructeurs AWS experts et à des exercices pratiques

Accédez à la bibliothèque d'apprentissage via le lien suivant : <https://aws.training>

# Merci!

Jonathan Dion  
Évangéliste technique principal  
Amazon Web Services

 @jotdion

 [linkedin.com/in/jotdion](https://www.linkedin.com/in/jotdion)

 SUMMIT

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.