

AWS Livre blanc

# AWS Outposts Considérations relatives à la conception et à l'architecture de haute disponibilité



---

# AWS Outposts Considérations relatives à la conception et à l'architecture de haute disponibilité: AWS Livre blanc

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Les marques et la présentation commerciale d'Amazon ne peuvent être utilisées en relation avec un produit ou un service qui n'est pas d'Amazon, d'une manière susceptible de créer une confusion parmi les clients, ou d'une manière qui dénigre ou discrédite Amazon. Toutes les autres marques commerciales qui ne sont pas la propriété d'Amazon appartiennent à leurs propriétaires respectifs, qui peuvent ou non être affiliés ou connectés à Amazon, ou sponsorisés par Amazon.

# Table of Contents

Résumé et introduction .....	i
Êtes-vous Well-Architected ? .....	1
Introduction .....	1
Extension de AWS l'infrastructure et des services aux sites sur site .....	2
Comprendre le modèle de responsabilité partagée actualisé .....	5
Penser en termes de modes de défaillance .....	7
Mode de défaillance 1 : réseau .....	7
Mode de défaillance 2 : Instances .....	8
Mode de défaillance 3 : calcul .....	8
Mode de défaillance 4 : racks ou centres de données .....	8
Mode de défaillance 5 : zone de AWS disponibilité ou région .....	9
Création d'applications HA et de solutions d'infrastructure avec un AWS Outposts rack .....	10
Réseaux .....	11
Rattachement au réseau .....	12
Connectivité d'ancrage .....	16
Routage des applications et des charges .....	20
Calcul .....	24
Planning des capacités .....	24
Gestion de capacité .....	28
Placement de l'instance .....	29
Stockage .....	32
Protection des données .....	33
Modes de défaillance plus importants .....	35
Conclusion .....	39
Collaborateurs .....	40
Historique de la documentation .....	41
Avis .....	42
AWS Glossaire .....	43
.....	xliv

# AWS Outposts Considérations relatives à la conception et à l'architecture de haute disponibilité

Date de publication : 12 août 2021 ([Historique de la documentation](#))

Ce livre blanc aborde les considérations relatives à l'architecture et les pratiques recommandées que les responsables informatiques et les architectes système peuvent appliquer pour créer des environnements d'applications sur site à haute disponibilité. AWS Outposts

## Êtes-vous Well-Architected ?

Le [AWS Well-Architected](#) Framework vous aide à comprendre les avantages et les inconvénients des décisions que vous prenez lors de la création de systèmes dans le cloud. Les six piliers du Framework vous permettent d'apprendre les meilleures pratiques architecturales pour concevoir et exploiter des systèmes fiables, sécurisés, efficaces, rentables et durables. À l'aide du [AWS Well-Architected Tool](#), disponible gratuitement dans le [AWS Management Console](#), vous pouvez évaluer votre charge de travail par rapport à ces meilleures pratiques en répondant à une série de questions pour chaque pilier.

[Pour obtenir des conseils d'experts supplémentaires et les meilleures pratiques relatives à votre architecture cloud \(déploiements d'architecture de référence, diagrammes et livres blancs\), consultez le Centre d'architecture.AWS](#)

## Introduction

Ce paper est destiné aux responsables informatiques et aux architectes système qui souhaitent déployer, migrer et exploiter des applications à l'aide de la plateforme AWS cloud et exécuter ces applications sur site avec un [AWS Outposts rack](#), le format rack 42U de [AWS Outposts](#).

Il présente les modèles d'architecture, les anti-modèles et les pratiques recommandées pour créer des systèmes hautement disponibles incluant des AWS Outposts racks. Vous apprendrez à gérer la capacité de vos AWS Outposts racks et à utiliser les services de mise en réseau et de centre de données pour configurer des solutions d'infrastructure de AWS Outposts rack à haute disponibilité.

AWS Outposts rack est un service entièrement géré qui fournit un pool logique de capacités de calcul, de stockage et de mise en réseau dans le cloud. [Avec les racks Outposts, les clients peuvent utiliser les services AWS gérés pris en charge dans leurs environnements sur site, notamment :](#)

[Amazon Elastic Compute Cloud \(Amazon EC2\)](#), [AmazonElastic Block Store \(Amazon EBS\)](#), [Amazon S3 on Outposts](#), [Amazon Elastic Kubernetes Service \(Amazon EKS\)](#), [Amazon Elastic Container Service \(Amazon ECS\)](#), [Amazon Relational Database Service Relational Database \(Amazon RDS\) et d'autres services sur Outposts.AWS](#) Les services sur les Outposts sont fournis sur le même [système AWS Nitro](#) que celui utilisé dans [le](#). Régions AWS

En tirant parti du AWS Outposts rack, vous pouvez créer, gérer et faire évoluer des applications sur site hautement disponibles à l'aide de services et d'outils AWS cloud familiers. AWS Outposts le rack est idéal pour les charges de travail qui nécessitent un accès à faible latence aux systèmes sur site, le traitement local des données, la résidence des données et la migration d'applications avec des interdépendances entre les systèmes locaux.

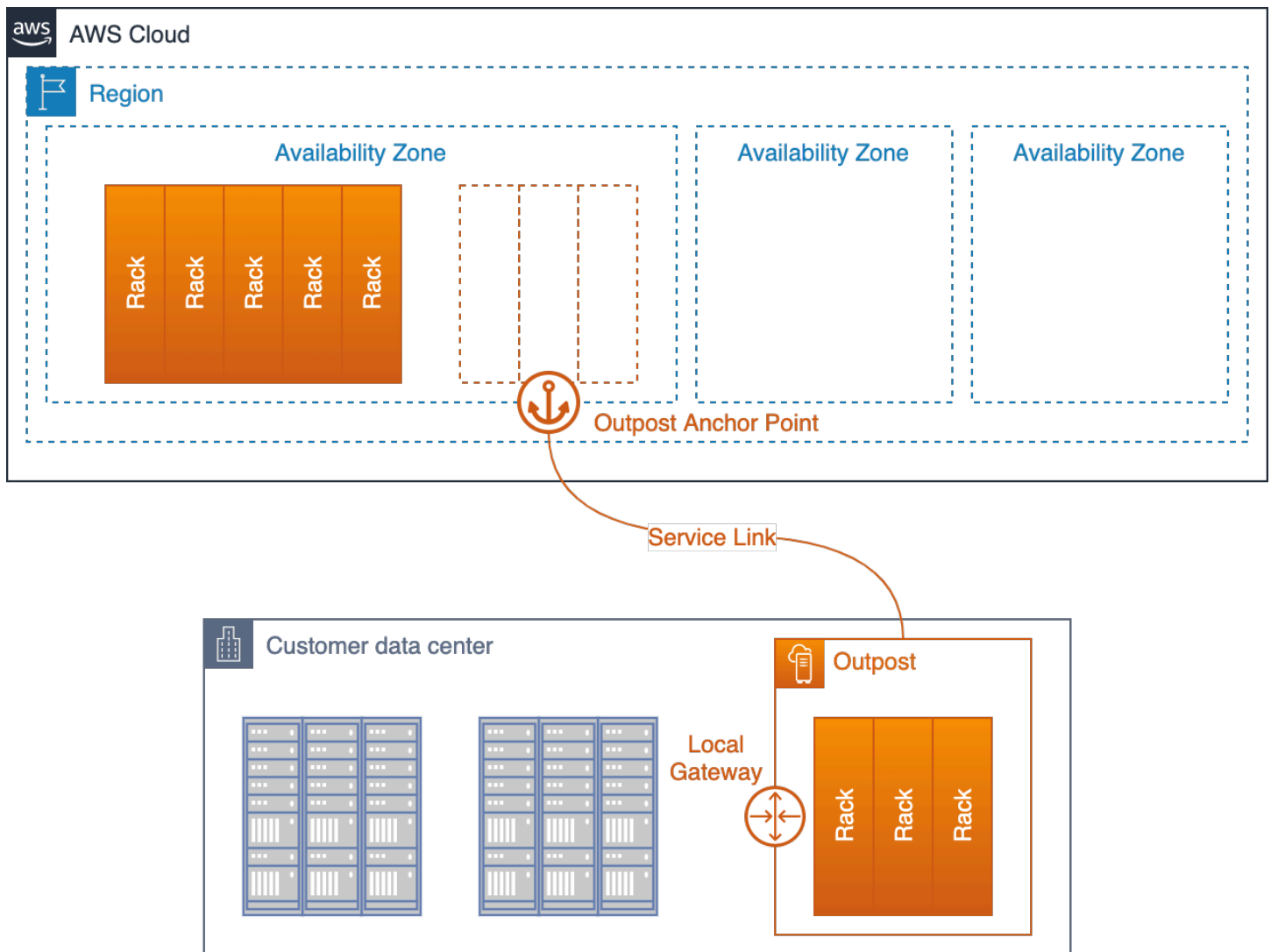
## Extension de AWS l'infrastructure et des services aux sites sur site

Le AWS Outposts service fournit une AWS infrastructure et des services à des sites sur site dans [plus de 50 pays et territoires](#), donnant aux clients la possibilité de déployer la même AWS infrastructure, les mêmes AWS services, les mêmes API et les mêmes outils dans pratiquement n'importe quel centre de données, espace de colocation ou installation sur site pour une expérience hybride véritablement cohérente. Pour comprendre comment concevoir avec Outposts, vous devez connaître les différents niveaux qui constituent le AWS cloud.

An [Région AWS](#) est une zone géographique du monde. Chacun Région AWS est un ensemble de centres de données regroupés logiquement en [zones de disponibilité](#) (AZ). Régions AWS fournir plusieurs (au moins deux) zones de disponibilité physiquement séparées et isolées qui sont connectées avec une faible latence, un débit élevé et une connectivité réseau redondante. Chaque AZ comprend un ou plusieurs centres de données physiques.

Un [avant-poste](#) logique (ci-après dénommé avant-poste) est un déploiement d'un ou de plusieurs AWS Outposts racks connectés physiquement gérés comme une seule entité. Un Outpost fournit un pool de capacités de AWS calcul et de stockage sur l'un de vos sites en tant qu'extension privée d'un AZ dans un Région AWS.

Le meilleur modèle conceptuel AWS Outposts est peut-être de penser à débrancher un ou plusieurs racks d'un centre de données situé dans un AZ ou un. Région AWS Vous faites rouler les racks du centre de données AZ vers votre centre de données. Vous branchez ensuite les racks aux points d'ancrage du centre de données AZ à l'aide d'un (très) long câble afin que les racks continuent de fonctionner en tant que partie intégrante du. Région AWS Vous les connectez également à votre réseau local pour fournir une connectivité à faible latence entre vos réseaux locaux et les charges de travail exécutées sur ces racks.



Un avant-poste déployé dans le centre de données d'un client et reconnecté à sa région d'origine (AZ) et à sa région mère

L'Outpost fonctionne comme une extension de l'AZ où il est ancré. AWS exploite, surveille et gère AWS Outposts l'infrastructure dans le cadre du Région AWS. Au lieu d'un très long câble physique, un Outpost se reconnecte à sa région mère via un ensemble de tunnels VPN cryptés appelés Service Link.

Le lien de service se termine sur un ensemble de points d'ancrage dans une zone de disponibilité (AZ) de la région mère de l'avant-poste.

C'est vous qui choisissez l'endroit où votre contenu est stocké. Vous pouvez répliquer et sauvegarder votre contenu à cet emplacement Région AWS ou à un autre emplacement. Votre contenu ne sera pas déplacé ou copié en dehors des emplacements que vous avez choisis sans votre accord,

sauf si cela est nécessaire pour se conformer à la loi ou à un ordre contraignant d'un organisme gouvernemental. Pour de plus amples informations, veuillez consulter [AWS FAQ sur la confidentialité des données](#).

Les charges de travail que vous déployez sur ces racks s'exécutent localement. Et, bien que la capacité de calcul et de stockage disponible dans ces racks soit limitée et ne puisse pas permettre d'exécuter les services cloud d'une Région AWS, les ressources déployées sur le rack (vos instances et leur stockage local) bénéficient des avantages d'une exécution locale alors que le plan de gestion continue de fonctionner dans la Région AWS.

Pour déployer des charges de travail sur un Outpost, vous ajoutez des sous-réseaux à vos environnements Virtual Private Cloud (VPC) et vous spécifiez un Outpost comme emplacement pour les sous-réseaux. Ensuite, vous sélectionnez le sous-réseau souhaité lors du déploiement des AWS ressources prises en charge par le biais de la CLI AWS Management Console, des API, du CDK ou des outils d'infrastructure en tant que code (IaC). Les instances des sous-réseaux Outpost communiquent avec d'autres instances de l'Outpost ou de la région via un réseau VPC.

L'Outpost Service Link transporte à la fois le trafic de gestion de l'Outpost et le trafic VPC du client (trafic VPC entre les sous-réseaux de l'Outpost et les sous-réseaux de la région).

### Termes importants :

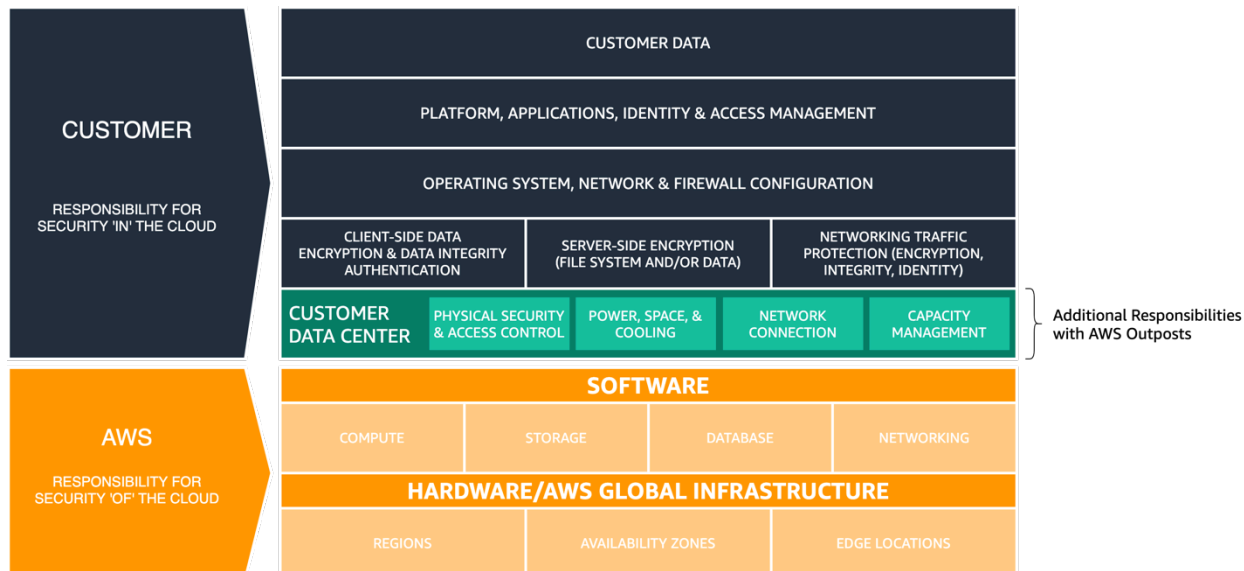
- **AWS Outposts**— est un service entièrement géré qui offre la même AWS infrastructure, les mêmes AWS services, les mêmes API et les mêmes outils à pratiquement tous les centres de données, espaces de colocation ou installations sur site pour une expérience hybride véritablement cohérente.
- **Outpost** : déploiement d'un ou de plusieurs AWS Outposts racks physiquement connectés gérés comme une entité logique unique et un pool de AWS calcul, de stockage et de mise en réseau déployés sur le site d'un client.
- **Région parent** : Région AWS qui fournit la gestion, les services de plan de contrôle et les AWS services régionaux pour le déploiement d'un avant-poste.
- **Zone de disponibilité des points d'ancrage (ancrage AZ)** : zone de disponibilité de la région parent qui héberge les points d'ancrage d'un avant-poste. Un avant-poste fonctionne comme une extension de sa zone de disponibilité principale.
- **Points d'ancrage** : points de terminaison situés dans la zone d'ancrage qui reçoivent les connexions des Outposts déployés à distance.
- **Lien de service** : ensemble de tunnels VPN cryptés qui connectent un avant-poste à sa zone de disponibilité principale dans sa région mère.

- Passerelle locale (LGW) : routeur virtuel d'interconnexion logique qui permet la communication entre votre Outpost et votre réseau local.

## Comprendre le modèle de responsabilité partagée actualisé

Lorsque vous déployez une AWS Outposts infrastructure dans vos centres de données ou vos installations de colocation, vous assumez des responsabilités supplémentaires dans le cadre du [modèle de responsabilité AWS partagée](#). Par exemple, dans la région, AWS fournit diverses sources d'alimentation, un réseau central redondant et une connectivité réseau étendu (WAN) résiliente pour garantir la disponibilité des services en cas de défaillance d'un ou de plusieurs composants.

Avec Outposts, vous êtes chargé de fournir une alimentation résiliente et une connectivité réseau aux racks Outpost afin de répondre à vos exigences de disponibilité pour les charges de travail exécutées sur les Outposts.



### AWS Modèle de responsabilité partagée mis à jour pour AWS Outposts

Avec AWS Outposts, vous êtes responsable de la sécurité physique et des contrôles d'accès de l'environnement du centre de données. Vous devez fournir suffisamment d'alimentation, d'espace et de refroidissement pour que l'avant-poste reste opérationnel et que les connexions réseau permettent de reconnecter l'avant-poste à la région.

La capacité des Outposts étant limitée et déterminée par la taille et le nombre de racks AWS installés sur votre site, vous devez décider de la capacité EC2, EBS et S3 on Outposts dont vous avez besoin pour exécuter vos charges de travail initiales, faire face à la croissance future et fournir une capacité supplémentaire afin d'atténuer les défaillances des serveurs et les événements de maintenance.



AWS est responsable de la disponibilité de l'infrastructure des Outposts, y compris les alimentations électriques, les serveurs et l'équipement réseau présents dans les AWS Outposts racks. AWS gère également l'hyperviseur de virtualisation, les systèmes de stockage et les AWS services exécutés sur Outposts.

Une étagère d'alimentation centrale située dans chaque rack Outposts convertit le courant alternatif en courant continu et alimente les serveurs du rack via une architecture de barre de bus. Grâce à l'architecture de barre de bus, la moitié des alimentations du rack peuvent tomber en panne et tous les serveurs continueront à fonctionner sans interruption.



Figure 3 : blocs d'alimentation AWS Outposts CA/CC et distribution de l'alimentation par barre omnibus

Les commutateurs réseau et le câblage à l'intérieur et entre les racks Outposts sont également totalement redondants. Un panneau de brassage en fibre optique assure la connectivité entre un rack Outpost et le réseau sur site et sert de point de démarcation entre l'environnement du centre de données géré par le client et l'environnement géré. AWS Outposts

Tout comme dans la région, AWS il est responsable des services cloud proposés sur Outposts et assume des responsabilités supplémentaires lorsque vous sélectionnez et déployez des services gérés de haut niveau tels qu'Amazon RDS on Outposts. Vous devriez consulter le [modèle de responsabilité AWS partagée](#) et les pages de questions fréquemment posées (FAQ) relatives aux différents services lorsque vous envisagez et sélectionnez les services à déployer sur les Outposts. Ces ressources fournissent des détails supplémentaires sur le partage des responsabilités entre vous et AWS.

# Penser en termes de modes de défaillance

Lorsque vous concevez une application ou un système à haute disponibilité, vous devez tenir compte des composants susceptibles de tomber en panne, de l'impact que les défaillances des composants auront sur le système et des mécanismes que vous pouvez mettre en œuvre pour atténuer ou éliminer l'impact des défaillances des composants. Votre application s'exécute-t-elle sur un seul serveur, dans un seul rack ou dans un seul centre de données ? Que se passera-t-il en cas de panne temporaire ou permanente d'un serveur, d'un rack ou d'un centre de données ? Que se passe-t-il en cas de défaillance d'un sous-système critique tel que le réseau ou de l'application elle-même ? Il s'agit de modes de défaillance.

Vous devez prendre en compte les modes de défaillance décrits dans cette section lors de la planification de vos Outposts et de vos déploiements d'applications. Les sections suivantes examinent comment atténuer ces modes de défaillance afin d'augmenter le niveau de haute disponibilité de votre environnement applicatif.

## Mode de défaillance 1 : réseau

Le déploiement d'un Outpost dépend d'une connexion résiliente à sa région mère à des fins de gestion et de surveillance. Les perturbations du réseau peuvent être causées par diverses défaillances, telles que des erreurs d'opérateur, des pannes d'équipement et des pannes de fournisseurs de services. Un avant-poste, qui peut être composé d'un ou de plusieurs racks connectés entre eux sur le site, est considéré comme déconnecté lorsqu'il ne peut pas communiquer avec la région via le lien de service.

Les chemins réseau redondants peuvent contribuer à atténuer le risque d'événements de déconnexion. Vous devez cartographier les dépendances des applications et le trafic réseau pour comprendre l'impact des événements de déconnexion sur les opérations de charge de travail. Prévoyez une redondance réseau suffisante pour répondre aux exigences de disponibilité de vos applications.

Lors d'un événement de déconnexion, les instances exécutées sur un avant-poste continuent de fonctionner et sont accessibles depuis les réseaux locaux via la passerelle locale de l'avant-poste (LGW). Les charges de travail et les services locaux peuvent être altérés ou échouer s'ils dépendent des services de la région. Les demandes mutantes (comme le démarrage ou l'arrêt d'instances sur l'avant-poste), les opérations du plan de contrôle et la télémétrie des services (par exemple, les CloudWatch métriques) échoueront lorsque l'avant-poste est déconnecté de la région.

## Mode de défaillance 2 : Instances

Les instances EC2 peuvent être altérées ou échouer si le serveur sur lequel elles s'exécutent rencontre un problème ou si l'instance rencontre une défaillance du système d'exploitation ou d'une application. La manière dont les applications gèrent ces types de défaillances dépend de l'architecture de l'application. Les applications monolithiques utilisent généralement les fonctionnalités de l'application ou du système pour la restauration, tandis que les architectures modulaires axées sur les services ou les microservices remplacent généralement les composants défaillants afin de maintenir la disponibilité des services.

Vous pouvez remplacer les instances défaillantes par de nouvelles instances à l'aide de mécanismes automatisés tels que les groupes EC2 Auto Scaling. La restauration automatique des instances peut redémarrer les instances qui échouent en raison de défaillances de serveur, à condition que la capacité de réserve disponible sur les serveurs restants soit suffisante.

## Mode de défaillance 3 : calcul

Les serveurs peuvent tomber en panne ou devenir défaillants et peuvent avoir besoin d'être mis hors service (temporairement ou définitivement) pour diverses raisons, telles que des défaillances de composants ou des opérations de maintenance planifiées. La manière dont les services du rack Outposts gèrent les défaillances et les défaillances des serveurs varie et peut dépendre de la manière dont les clients configurent les options de haute disponibilité.

Vous devez commander une capacité de calcul suffisante pour prendre en charge un modèle de N +M disponibilité, dans lequel N la capacité requise et M la capacité de réserve sont-elles allouées pour faire face aux défaillances des serveurs ?

Le remplacement du matériel pour les serveurs défaillants est fourni dans le cadre du service de AWS Outposts rack entièrement géré. AWS surveille activement l'état de tous les serveurs et équipements réseau dans le cadre d'un déploiement Outpost. S'il est nécessaire d'effectuer une maintenance physique, AWS planifiera une visite sur votre site afin de remplacer les composants défectueux. Le provisionnement en capacité de réserve vous permet de continuer à exécuter vos charges de travail pendant que les serveurs défaillants sont mis hors service et remplacés.

## Mode de défaillance 4 : racks ou centres de données

Les défaillances des racks peuvent être dues à une perte totale d'alimentation des racks ou à des défaillances environnementales telles qu'une perte de refroidissement ou des dommages physiques

au centre de données à la suite d'une inondation ou d'un tremblement de terre. Des défaillances dans les architectures de distribution électrique des centres de données ou des erreurs lors de la maintenance standard de l'alimentation des centres de données peuvent entraîner la perte d'alimentation d'un ou de plusieurs racks, voire de l'ensemble du centre de données.

Ces scénarios peuvent être atténués en déployant l'infrastructure sur plusieurs étages ou sites de centres de données indépendants les uns des autres au sein du même campus ou de la même région métropolitaine.

L'adoption de cette approche avec le AWS Outposts rack nécessitera une attention particulière à la manière dont les applications sont architecturées et distribuées pour fonctionner sur plusieurs Outposts logiques distincts afin de maintenir la disponibilité des applications.

## Mode de défaillance 5 : zone de AWS disponibilité ou région

Chaque avant-poste est ancré dans une zone de disponibilité (AZ) spécifique au sein d'un Région AWS. Les défaillances au sein de l'AZ d'ancrage ou de la région mère peuvent entraîner la perte de gestion et de mutabilité de l'avant-poste et perturber les communications réseau entre l'avant-poste et la région.

À l'instar des pannes de réseau, les défaillances de l'AZ ou de la région peuvent entraîner la déconnexion de l'avant-poste de la région. Les instances exécutées sur un avant-poste continuent de fonctionner et sont accessibles depuis les réseaux locaux via la passerelle locale de l'avant-poste (LGW) et peuvent être défectueuses ou échouer si elles dépendent des services de la région, comme décrit précédemment.

Pour atténuer l'impact des défaillances d' AWS une zone ou d'une région, vous pouvez déployer plusieurs Outposts, chacun ancré dans une zone ou une région différente. Vous pouvez ensuite concevoir votre charge de travail pour qu'elle fonctionne dans un modèle de déploiement distribué multi-avant-postes en utilisant de nombreux [mécanismes et modèles architecturaux](#) similaires à ceux que vous utilisez pour concevoir et déployer aujourd'hui. AWS

# Création d'applications HA et de solutions d'infrastructure avec un AWS Outposts rack

Avec le AWS Outposts rack, vous pouvez créer, gérer et faire évoluer des applications sur site hautement disponibles à l'aide de services et d'outils AWS cloud familiers. Il est important de comprendre que les architectures et approches de haute disponibilité dans le cloud sont généralement différentes des architectures HA sur site traditionnelles que vous utilisez peut-être aujourd'hui dans votre centre de données.

Dans le cadre des déploiements d'applications HA traditionnels sur site, les applications sont déployées sur des machines virtuelles (VM). Les systèmes et infrastructures informatiques complexes sont déployés et entretenus pour assurer le bon fonctionnement et le bon fonctionnement de ces machines virtuelles. Les machines virtuelles ont souvent des identités spécifiques et chaque machine virtuelle peut jouer un rôle essentiel dans l'architecture globale de l'application.

Les rôles architecturaux sont étroitement liés aux identités des machines virtuelles. Les architectes systèmes tirent parti des fonctionnalités de l'infrastructure informatique pour fournir des environnements d'exécution de machines virtuelles hautement disponibles qui fournissent à chaque machine virtuelle un accès fiable à la capacité de calcul, aux volumes de stockage et aux services réseau. En cas de défaillance d'une machine virtuelle, des processus de restauration automatisés ou manuels sont exécutés pour rétablir l'état sain de la machine virtuelle défaillante, souvent sur une autre infrastructure ou dans un autre centre de données entièrement différent.

Les architectures Cloud HA adoptent une approche différente. AWS les services cloud fournissent des capacités de calcul, de stockage et de mise en réseau fiables. Les composants de l'application sont déployés sur des instances EC2, des conteneurs, des fonctions sans serveur ou d'autres services gérés.

Une instance est une instanciation d'un composant d'application, peut-être l'une des nombreuses instances jouant ce rôle. Les composants de l'application sont faiblement couplés les uns aux autres et au rôle qu'ils jouent dans l'architecture globale de l'application. L'identité individuelle d'une instance n'est généralement pas importante. Des instances supplémentaires peuvent être créées ou détruites pour augmenter ou diminuer la taille en réponse à la demande. Les instances défaillantes ou défectueuses sont simplement remplacées par de nouvelles instances saines.

AWS Outposts rack est un service entièrement géré qui étend les services de AWS calcul, de stockage, de mise en réseau, de base de données et d'autres services cloud aux sites sur site pour

une expérience hybride véritablement cohérente. Vous ne devez pas considérer le service Outposts rack comme un remplacement direct des systèmes d'infrastructure informatique par des mécanismes de haute disponibilité sur site traditionnels. Tenter d'utiliser AWS des services et des Outposts pour prendre en charge une architecture HA traditionnelle sur site est un contre-modèle.

Les charges de travail exécutées sur AWS Outposts rack utilisent des mécanismes de haute disponibilité dans le cloud tels qu'[Amazon EC2 Auto Scaling](#) (pour effectuer une mise à l'échelle horizontale afin de répondre aux demandes de charge de travail), [les bilans de santé EC2](#) (pour détecter et supprimer les instances défectueuses) et les équilibrateurs de charge d'[application \(pour rediriger le trafic de charge](#) de travail entrant vers des instances redimensionnées ou remplacées). Lorsque vous migrez des applications vers le cloud, que ce soit vers un AWS Outposts rack Région AWS ou un rack, vous devez mettre à jour l'architecture de votre application HA pour commencer à tirer parti des services cloud gérés et des mécanismes de haute disponibilité dans le cloud.

Les sections suivantes présentent les modèles d'architecture, les anti-modèles et les pratiques recommandées pour déployer des AWS Outposts racks dans vos environnements sur site afin d'exécuter des charges de travail répondant à des exigences de haute disponibilité. Ces sections présentent des modèles et des pratiques ; toutefois, elles ne fournissent pas de détails sur la configuration et la mise en œuvre. Vous devriez lire et vous familiariser avec les [FAQ et le guide de l'utilisateur, ainsi que les FAQ et la documentation des services exécutés sur le AWS Outposts rack](#) Outposts pendant que vous préparez votre environnement pour le rack Outposts et vos applications pour la migration vers les services. AWS

## Rubriques

- [Réseaux](#)
- [Calcul](#)
- [Stockage](#)
- [Modes de défaillance plus importants](#)

## Réseaux

Le déploiement d'un Outpost dépend d'une connexion résiliente à son AZ d'ancrage pour que les opérations de gestion, de surveillance et de service fonctionnent correctement. Vous devez configurer votre réseau sur site de manière à fournir des connexions réseau redondantes pour chaque rack Outpost et une connectivité fiable vers les points d'ancrage dans le cloud. AWS Tenez également compte des chemins réseau entre les charges de travail des applications exécutées

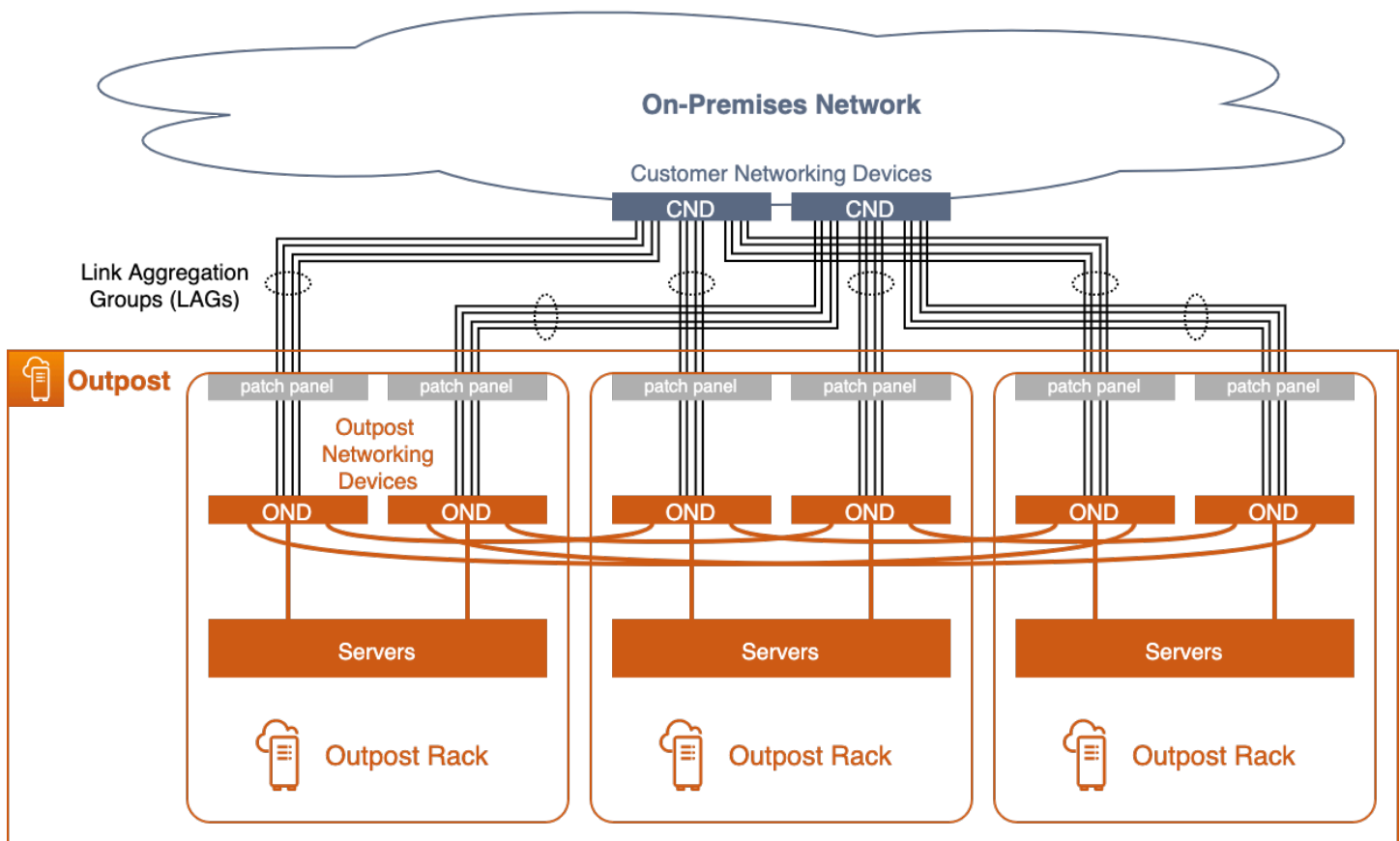
sur l'Outpost et les autres systèmes sur site et dans le cloud avec lesquels elles communiquent : comment acheminerez-vous ce trafic sur votre réseau ?

## Rubriques

- [Rattachement au réseau](#)
- [Connectivité d'ancrage](#)
- [Routage des applications et des charges](#)

## Rattachement au réseau

Chaque AWS Outposts rack est configuré avec des top-of-rack commutateurs redondants appelés Outpost Networking Devices (OND). Les serveurs de calcul et de stockage de chaque rack se connectent aux deux OND. Vous devez connecter chaque OND à un commutateur distinct appelé périphérique réseau client (CND) de votre centre de données afin de fournir divers chemins physiques et logiques pour chaque rack Outpost. Les OND se connectent à vos CND via une ou plusieurs connexions physiques à l'aide de câbles à fibres optiques et d'émetteurs-récepteurs optiques. Les [connexions physiques](#) sont configurées dans des [liens de groupes d'agrégation de liens logiques \(LAG\)](#).



### Avant-poste à plusieurs racks avec connexions réseau redondantes

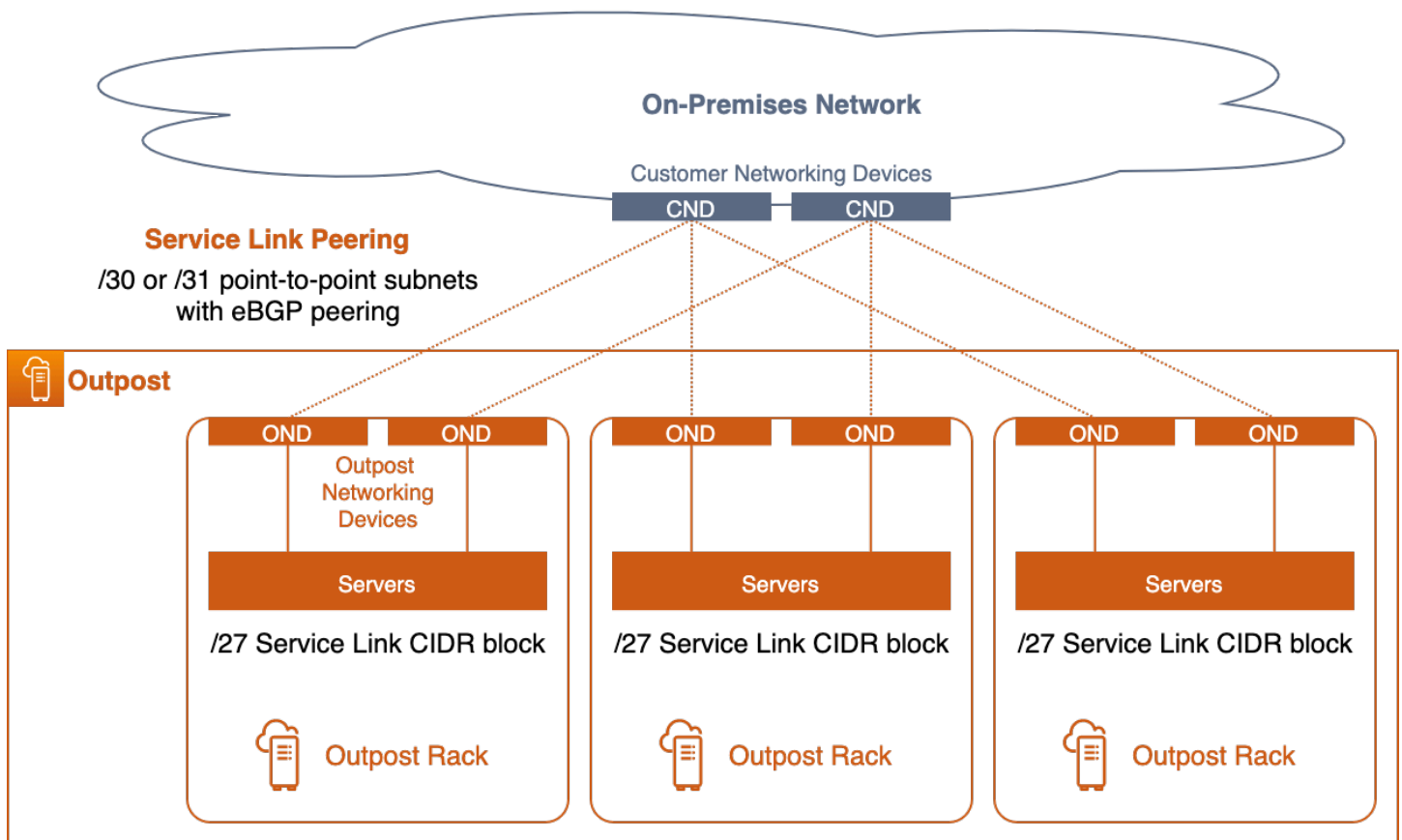
Les liaisons OND vers CND sont toujours configurées dans un LAG, même si la connexion physique est un seul câble à fibre optique. La configuration des liens en tant que groupes LAG vous permet d'augmenter la bande passante des liens en ajoutant des connexions physiques supplémentaires au groupe logique. Les liaisons LAG sont configurées comme des jonctions Ethernet IEEE 802.1q afin de permettre une mise en réseau séparée entre l'Outpost et le réseau local.

Chaque avant-poste possède au moins deux réseaux logiquement séparés qui doivent communiquer avec ou via le réseau du client :

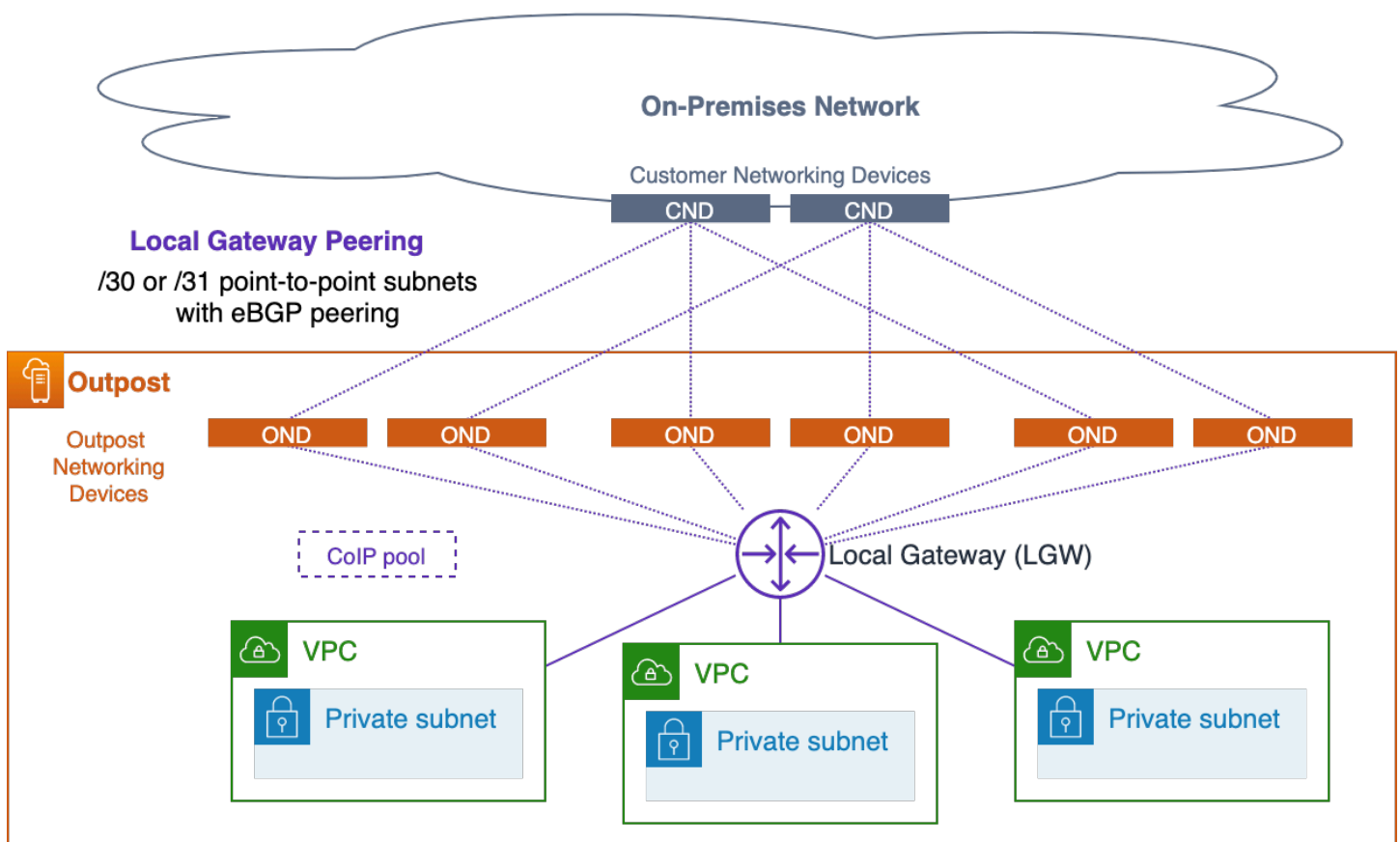
- Réseau de liaison de service : attribue les adresses IP des liaisons de service aux serveurs de l'avant-poste et facilite la communication avec le réseau local pour permettre aux serveurs de se reconnecter aux points d'ancrage de l'avant-poste dans la région.
- Réseau de passerelle local : permet la communication entre les sous-réseaux VPC de l'avant-poste et le réseau local via la passerelle locale de l'avant-poste (LGW).



Ces réseaux séparés se connectent au réseau local par un ensemble de [connexions point-to-point IP via les liaisons LAG](#). Chaque liaison LAG OND vers CND est configurée avec des identifiants VLAN, des sous-réseaux IP point-to-point (/30 ou /31) et un peering eBGP pour chaque réseau séparé (lien de service et LGW). Vous devez considérer les liens LAG, avec leurs point-to-point VLAN et leurs sous-réseaux, comme des connexions de couche 3 segmentées de couche 2 et routées. Les connexions IP routées fournissent des chemins logiques redondants qui facilitent la communication entre les réseaux séparés de l'Outpost et le réseau local.



### Peering des liens de service



### Peering via une passerelle locale

Vous devez mettre fin aux liaisons LAG de couche 2 (et à leurs VLAN) sur les commutateurs CND directement connectés et configurer les interfaces IP et le peering BGP sur les commutateurs CND. Vous ne devez pas relier les VLAN LAG entre les commutateurs de votre centre de données. Pour plus d'informations, consultez la section [Connectivité de la couche réseau](#) dans le guide de AWS Outposts l'utilisateur.

Au sein d'un avant-poste logique à plusieurs racks, les OND sont interconnectés de manière redondante pour fournir une connectivité réseau hautement disponible entre les racks et les charges de travail exécutées sur les serveurs. AWS est responsable de la disponibilité du réseau au sein de l'avant-poste.

### Pratiques recommandées pour une connexion réseau à haute disponibilité

- Connectez chaque périphérique réseau Outpost (OND) d'un rack Outpost à un périphérique réseau client (CND) distinct du centre de données.

- Mettez fin aux liaisons de couche 2, aux VLAN, aux sous-réseaux IP de couche 3 et au peering BGP sur les commutateurs CND (Customer Networking Device) directement connectés. Ne reliez pas les VLAN OND aux VLAN CND entre les CND ou sur le réseau local.
- Ajoutez des liens vers les groupes d'agrégation de liens (LAG) pour augmenter la bande passante disponible entre l'avant-poste et le centre de données. Ne vous fiez pas à la bande passante agrégée des différents chemins passant par les deux OND.
- Utilisez les différents chemins via les ONS redondants pour fournir une connectivité résiliente entre les réseaux Outpost et le réseau sur site.
- Pour obtenir une redondance optimale et permettre une maintenance OND sans interruption, nous recommandons aux clients de configurer les publicités et les politiques BGP comme suit :
  - L'équipement réseau du client doit recevoir des publicités BGP d'Outpost sans modifier les attributs BGP et pour permettre l'équilibrage de charge et les trajets multiples BGP afin d'obtenir des flux de trafic entrant optimaux (du client vers Outpost). Le préfixe AS-Path est utilisé pour les préfixes BGP d'Outpost afin de déplacer le trafic vers un autre OND/Uplink au cas où une maintenance serait requise. Le réseau client doit préférer les itinéraires depuis Outpost avec un chemin AS-Path de longueur 1 aux itinéraires avec un chemin AS de longueur 4, c'est-à-dire réagir au préfixe AS-Path.
  - Le réseau client doit annoncer des préfixes BGP identiques avec les mêmes attributs pour tous les OND d'Outpost. Par défaut, le réseau Outpost équilibre la charge du trafic sortant (vers le client) entre toutes les liaisons montantes. Les politiques de routage sont utilisées du côté de l'avant-poste pour détourner le trafic d'un OND particulier au cas où une maintenance serait requise. Des préfixes BGP identiques du côté client sur toutes les OND sont nécessaires pour effectuer ce transfert de trafic et effectuer la maintenance de manière non perturbatrice. Lorsqu'une maintenance est requise sur le réseau du client, nous recommandons d'utiliser le préfixe AS-Path pour éloigner temporairement le trafic d'une liaison montante ou d'un appareil en particulier.

## Connectivité d'ancrage

Un [lien de service Outpost](#) se connecte à des points d'ancrage publics ou privés (pas les deux) dans une zone de disponibilité (AZ) spécifique de la région parent de l'Outpost. Les serveurs Outpost initient des connexions VPN sortantes à partir de leurs adresses IP de liaison de service vers les points d'ancrage de l'AZ d'ancrage. Ces connexions utilisent les ports UDP et TCP 443. AWS est responsable de la disponibilité des points d'ancrage dans la Région.

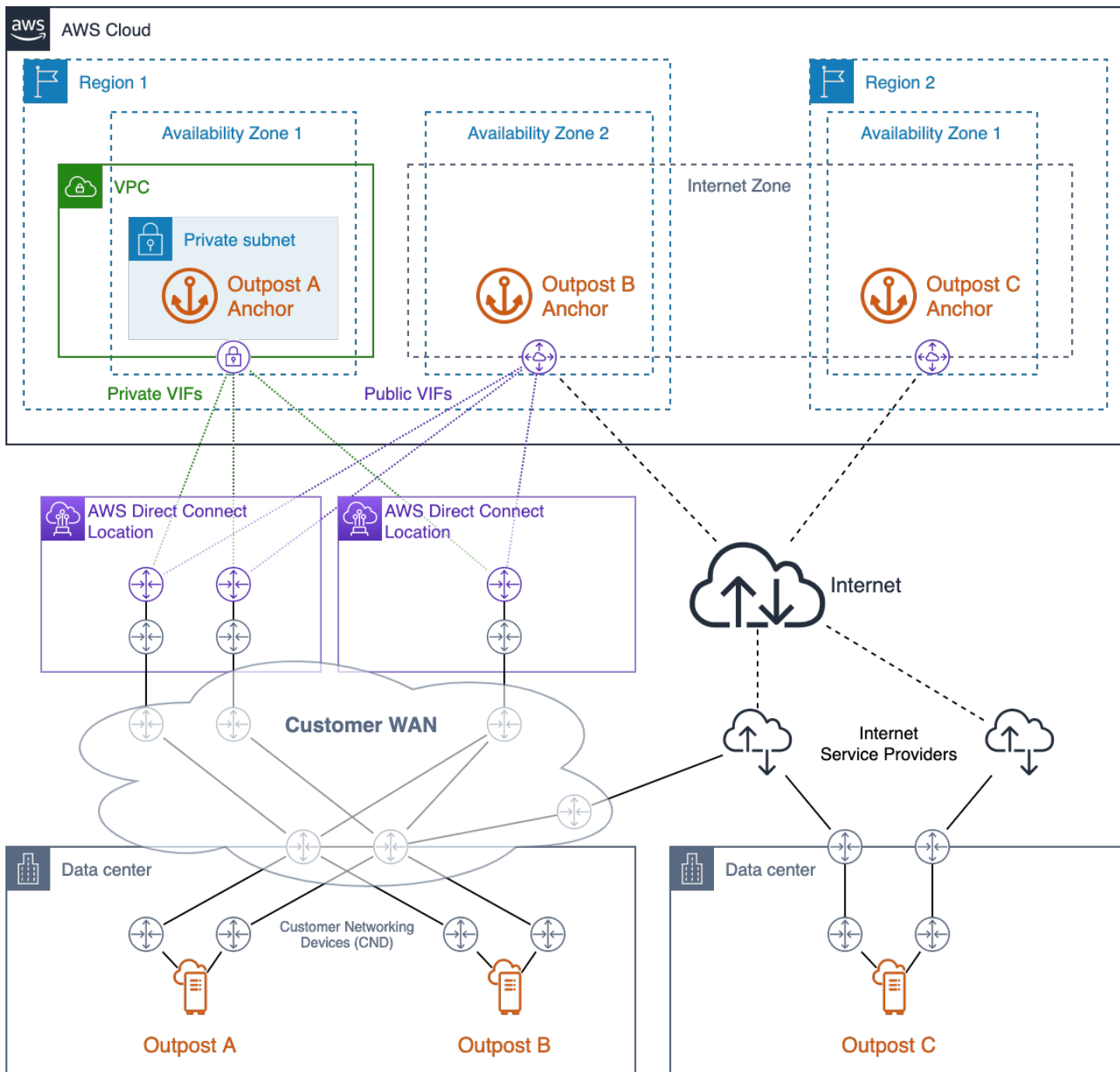
Vous devez vous assurer que les adresses IP du lien du service Outpost peuvent être connectées via votre réseau aux points d'ancrage de l'AZ d'ancrage. Les adresses IP du lien de service n'ont pas besoin de communiquer avec les autres hôtes de votre réseau local.

Les points d'ancrage publics se trouvent dans les [plages d'adresses IP publiques](#) de la région (dans les blocs CIDR du service EC2) et sont accessibles via Internet ou [AWS Direct Connect](#)(DX) des interfaces virtuelles publiques (VIF). L'utilisation de points d'ancrage publics permet une sélection plus flexible du chemin, car le trafic des liaisons de service peut être acheminé sur n'importe quel chemin disponible pouvant atteindre avec succès les points d'ancrage sur l'Internet public.

Les points d'ancrage privés vous permettent d'utiliser vos plages d'adresses IP pour la connectivité d'ancrage. Les points d'ancrage privés sont créés dans un [sous-réseau privé au sein d'un VPC dédié à](#) l'aide d'adresses IP attribuées par le client. Le VPC est créé dans le propriétaire de la ressource Outpost et vous êtes chargé de vous assurer Compte AWS que le VPC est disponible et correctement configuré (ne le supprimez pas !). Les points d'ancrage privés doivent être accessibles à l'aide [des VIF privés Direct Connect](#).

Vous devez fournir des chemins réseau redondants entre l'avant-poste et les points d'ancrage de la région, les connexions se terminant sur des appareils distincts situés à plusieurs endroits. Le routage dynamique doit être configuré pour rediriger automatiquement le trafic vers des chemins alternatifs en cas de défaillance des connexions ou des périphériques réseau. Vous devez prévoir une capacité réseau suffisante pour garantir que la défaillance d'un chemin WAN ne submerge pas les chemins restants.

Le schéma suivant montre trois Outposts dotés de chemins réseau redondants vers leurs zones d'ancrage, AWS Direct Connect ainsi que d'une connectivité Internet publique. L'avant-poste A et l'avant-poste B sont ancrés dans différentes zones de disponibilité dans la même région. L'avant-poste A se connecte aux points d'ancrage privés de l'AZ 1 de la région 1. L'avant-poste B se connecte aux points d'ancrage publics de l'AZ 2 de la région 1. L'avant-poste C se connecte aux points d'ancrage publics de l'AZ 1 de la région 2.



### Connectivité d'ancrage hautement disponible avec AWS Direct Connect accès public à Internet

L'avant-poste A dispose de trois chemins réseau redondants pour atteindre son point d'ancrage privé. Deux voies sont disponibles via des circuits Direct Connect redondants situés sur un seul emplacement Direct Connect. Le troisième chemin est disponible via un circuit Direct Connect à un deuxième emplacement Direct Connect. Cette conception permet de maintenir le trafic des liaisons de service de l'Outpost A sur les réseaux privés et assure la redondance des chemins, ce qui permet

la défaillance de l'un des circuits Direct Connect ou la défaillance d'un emplacement Direct Connect complet.

L'avant-poste B dispose de quatre chemins réseau redondants pour atteindre son point d'ancrage public. Trois chemins sont disponibles via des VIF publics fournis sur les circuits Direct Connect et les emplacements utilisés par Outpost A. Le quatrième chemin est disponible via le WAN du client et l'Internet public. Le trafic des liaisons de service de l'Outpost B peut être acheminé par n'importe quel chemin disponible permettant d'atteindre les points d'ancrage sur l'Internet public. L'utilisation des chemins Direct Connect peut fournir une latence plus constante et une meilleure disponibilité de bande passante, tandis que le chemin Internet public peut être utilisé pour des scénarios de reprise après sinistre (DR) ou d'augmentation de bande passante.

L'Outpost C dispose de deux chemins réseau redondants pour atteindre son point d'ancrage public. L'Outpost C est déployé dans un centre de données différent de celui des Outposts A et B. Le centre de données de l'Outpost C ne dispose pas de circuits dédiés connectés au WAN du client. Au lieu de cela, le centre de données dispose de connexions Internet redondantes fournies par deux fournisseurs de services Internet (ISP) différents. Le trafic des liaisons de service d'Outpost C peut être acheminé via l'un des réseaux ISP pour atteindre les points d'ancrage sur l'Internet public. Cette conception offre la flexibilité nécessaire pour acheminer le trafic des liaisons de service sur n'importe quelle connexion Internet publique disponible. Cependant, le end-to-end chemin dépend des réseaux tiers publics où la disponibilité de la bande passante et la latence du réseau fluctuent.

Le chemin réseau entre un avant-poste et ses points d'ancrage de liaison de service doit respecter les spécifications de bande passante suivantes :

- 500 Mbits/s à 1 Gbit/s de bande passante disponible par rack Outpost (par exemple, 3 racks : bande passante disponible de 1,5 à 3 Gbit/s)

Pratiques recommandées pour une connectivité d'ancrage à haute disponibilité :

- Fournissez des chemins réseau redondants entre chaque avant-poste et ses points d'ancrage dans la région.
- Utilisez les chemins Direct Connect (DX) pour contrôler la latence et la disponibilité de la bande passante.
- Assurez-vous que les ports TCP et UDP 443 sont ouverts (sortants) entre les blocs CIDR Outpost Service Link et les plages d'[adresses IP EC2](#) de la région parent. Assurez-vous que les ports sont ouverts sur tous les chemins réseau.

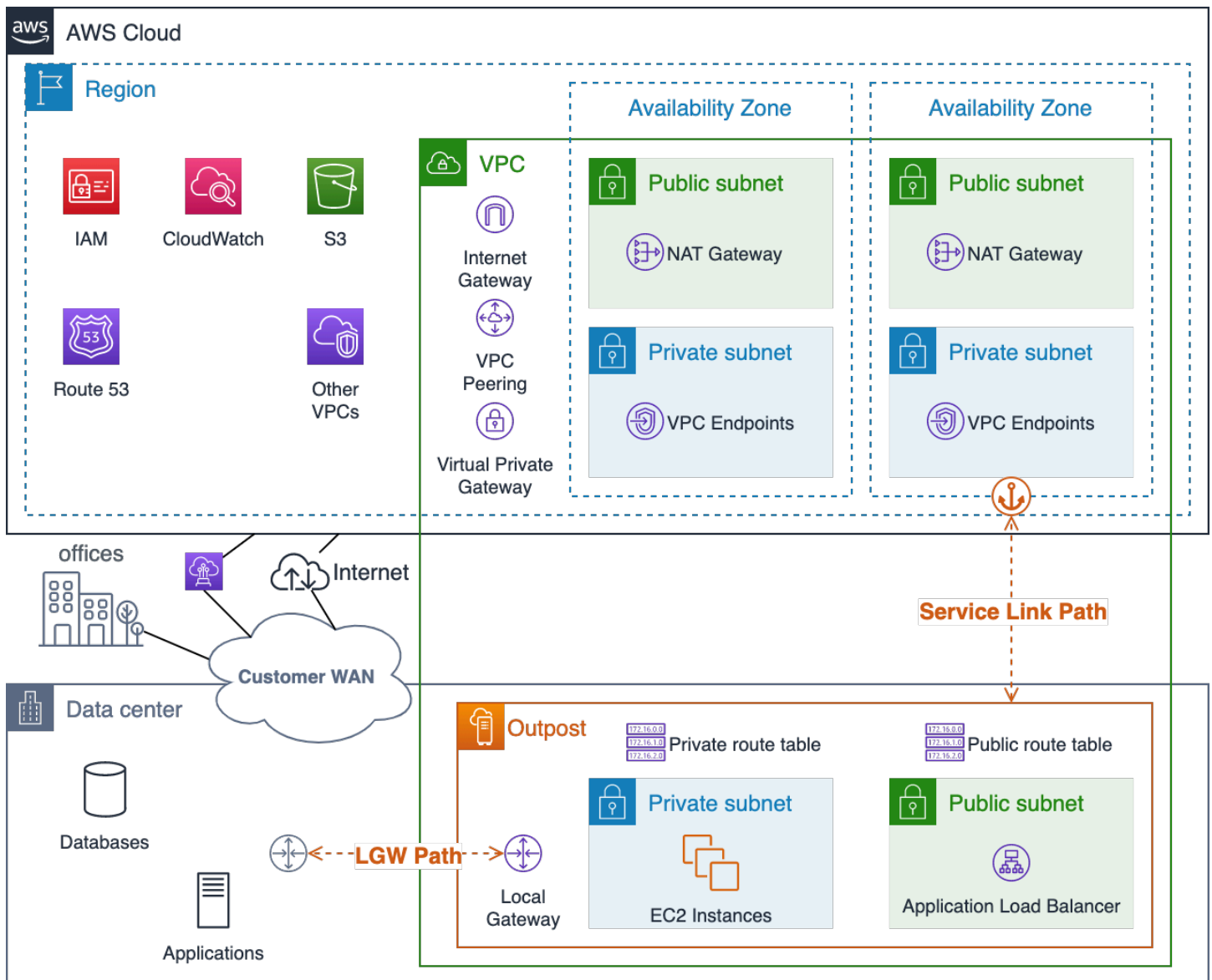
- Assurez-vous que chaque chemin répond aux exigences de disponibilité de bande passante et de latence.
- Utilisez le routage dynamique pour automatiser la redirection du trafic en cas de défaillance du réseau.
- Testez le routage du trafic de liaison de service sur chaque chemin réseau planifié pour vous assurer que le chemin fonctionne comme prévu.

## Routage des applications et des charges

Il existe deux voies pour sortir de l'Outpost pour les charges de travail des applications :

- Le chemin du lien de service
- Le chemin de la passerelle locale (LGW)

Vous configurez les tables de routage du sous-réseau Outpost pour contrôler le chemin à emprunter pour atteindre les réseaux de destination. Les itinéraires pointés vers le LGW dirigeront le trafic depuis la passerelle locale vers le réseau local. Les itinéraires pointant vers les services et ressources de la région, tels que Internet Gateway, NAT Gateway, Virtual Private Gateway et TGW, utiliseront le [lien de service](#) pour atteindre ces cibles. Si vous disposez d'une connexion d'appairage VPC avec plusieurs VPC sur le même avant-poste, le trafic entre les VPC reste sur l'avant-poste et n'utilise pas le lien de service vers la région. Pour plus d'informations sur le peering VPC, consultez [Connect VPC using VPC peering dans le guide de l'utilisateur Amazon VPC](#).



## Visualisation du lien de service Outpost et des chemins du réseau LGW

Lorsque vous planifiez le routage des applications, veillez à prendre en compte à la fois le fonctionnement normal et le routage limité et la disponibilité des services en cas de défaillance du réseau. Le chemin du lien de service n'est pas disponible lorsqu'un avant-poste est déconnecté de la région.

Vous devez prévoir différents chemins et configurer le routage dynamique entre l'Outpost LGW et vos applications, systèmes et utilisateurs critiques sur site. Les chemins réseau redondants permettent au réseau d'acheminer le trafic en cas de panne et de garantir que les ressources locales seront en mesure de communiquer avec les charges de travail exécutées sur l'avant-poste en cas de défaillance partielle du réseau.



Les configurations de routage VPC d'Outpost sont statiques. Vous configurez les tables de routage de sous-réseau par le biais de la CLI AWS Management Console, des API et d'autres outils d'infrastructure en tant que code (IaC) ; toutefois, vous ne pourrez pas modifier les tables de routage de sous-réseau lors d'un événement de déconnexion. Vous devrez rétablir la connectivité entre l'avant-poste et la région pour mettre à jour les tables de routage. Pour les opérations normales, utilisez les mêmes itinéraires que ceux que vous prévoyez d'utiliser lors d'événements de déconnexion.

Les ressources de l'Outpost peuvent accéder à Internet via le lien de service et une passerelle Internet (IGW) dans la région ou via le chemin de la passerelle locale (LGW). Le routage du trafic Internet via le chemin LGW et le réseau local vous permet d'utiliser les points d'entrée/sortie Internet existants sur site et peut fournir une latence plus faible, des MTU plus élevés et des frais de sortie de AWS données réduits par rapport à l'utilisation du chemin de liaison de service vers un IGW de la région.

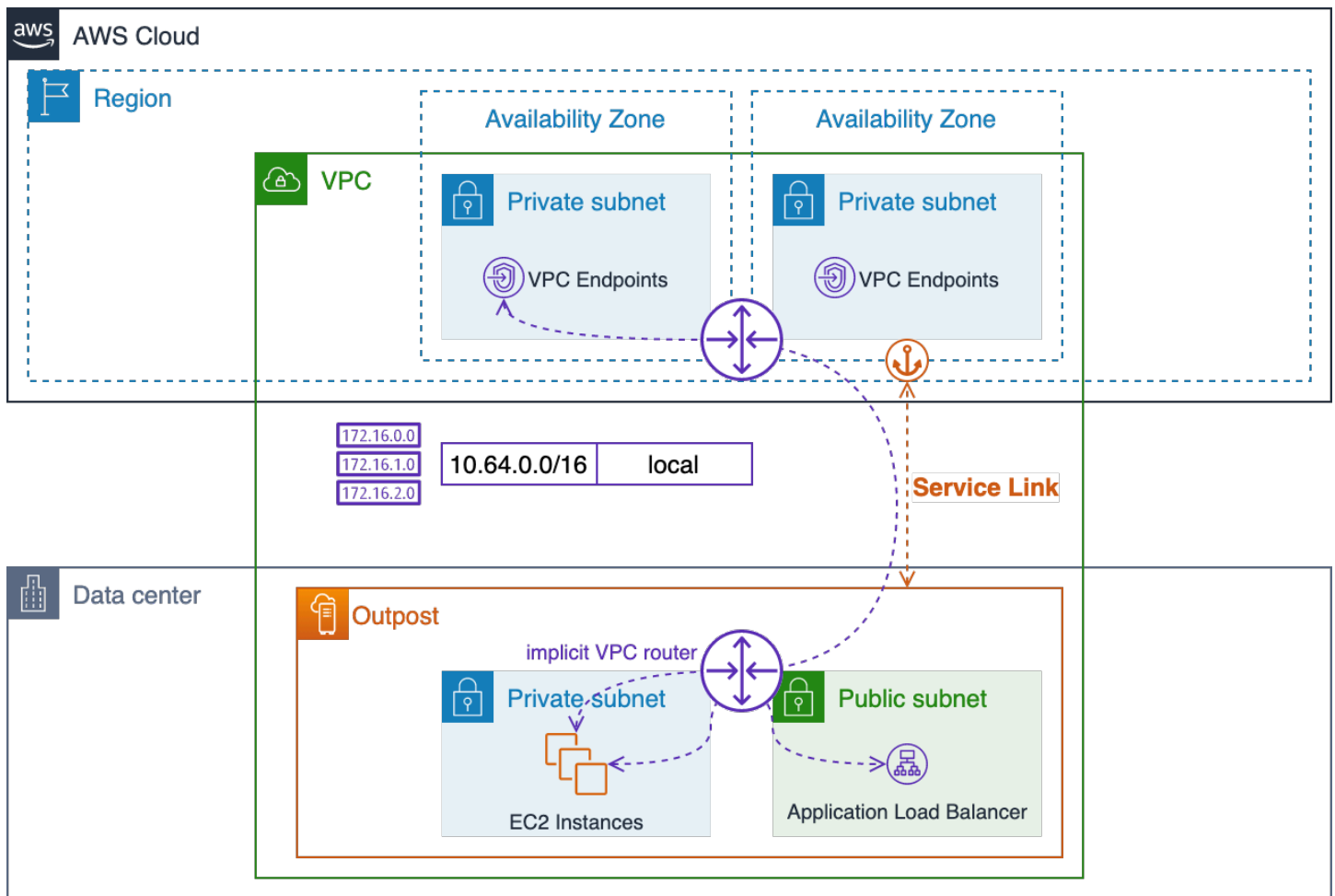
Si votre application doit s'exécuter sur site et être accessible depuis l'Internet public, vous devez acheminer le trafic de l'application via vos connexions Internet locales vers le LGW pour atteindre les ressources de l'Outpost.

Bien que vous puissiez configurer des sous-réseaux sur un avant-poste, comme les sous-réseaux publics de la région, cette pratique peut s'avérer indésirable dans la plupart des cas d'utilisation. Le trafic Internet entrant entrera par le lien de service Région AWS et sera acheminé vers les ressources exécutées sur l'avant-poste.

Le trafic de réponse sera à son tour acheminé via le lien de service et renvoyé via les connexions Internet Région AWS de celui-ci. Ce schéma de trafic peut ajouter de la latence et entraîner des frais de sortie de données lorsque le trafic quitte la région pour se rendre à l'avant-poste et lorsque le trafic de retour revient par la région et sort vers Internet. Si votre application peut être exécutée dans la région, la région est le meilleur endroit pour l'exécuter.

Le trafic entre les ressources VPC (dans le même VPC) suivra toujours la route CIDR VPC locale et sera acheminé entre les sous-réseaux par les routeurs VPC implicites.

Par exemple, le trafic entre une instance EC2 exécutée sur l'Outpost et un point de terminaison VPC dans la région sera toujours acheminé via le lien de service.



### Routing VPC local via les routeurs implicites

### Pratiques recommandées pour le routage des applications et des charges de travail :

- Utilisez le chemin de la passerelle locale (LGW) au lieu du chemin du lien de service dans la mesure du possible.
- Acheminez le trafic Internet sur le chemin LGW.
- Configurez les tables de routage du sous-réseau Outpost avec un ensemble standard de routes. Elles seront utilisées à la fois pour les opérations normales et lors des événements de déconnexion.
- Fournissez des chemins réseau redondants entre l'Outpost LGW et les ressources applicatives critiques sur site. Utilisez le routage dynamique pour automatiser la redirection du trafic en cas de défaillance du réseau sur site.

# Calcul

Alors que la capacité d'Amazon EC2 Régions AWS est apparemment infinie, la capacité des Outposts est limitée. Vous êtes responsable de la planification et de la gestion de la capacité de calcul de vos déploiements d'Outposts.

## Rubriques

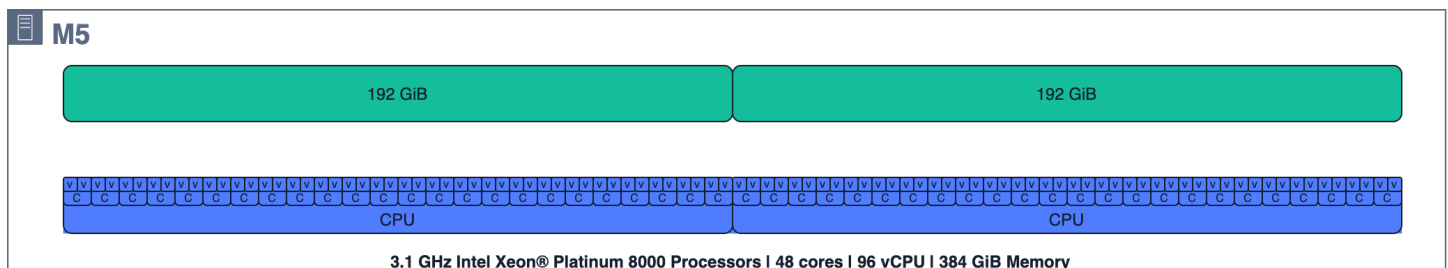
- [Planning des capacités](#)
- [Gestion de capacité](#)
- [Placement de l'instance](#)

## Planning des capacités

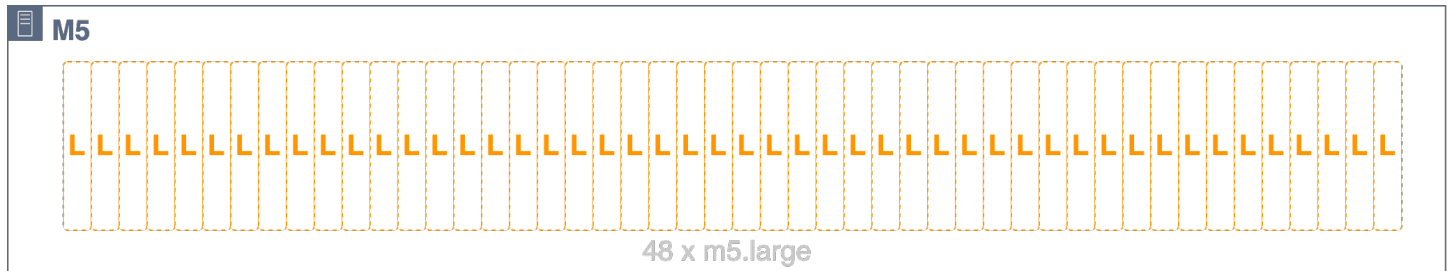
Alors que la capacité entrante d'Amazon EC2 Régions AWS est apparemment infinie, la capacité des Outposts est limitée, limitée par le volume total de capacité de calcul commandé. Vous êtes responsable de la planification et de la gestion de la capacité de calcul de vos déploiements d'Outposts. Vous devez commander une capacité de calcul suffisante pour prendre en charge un modèle de disponibilité N+M, dans lequel N est le nombre de serveurs requis et M le nombre de serveurs de rechange fournis pour faire face aux défaillances des serveurs. N+1 et N+2 sont les niveaux de disponibilité les plus courants.

Chaque serveur (C5,, M5R5, etc.) prend en charge une seule famille d'instances EC2. Avant de pouvoir lancer des instances sur des serveurs de calcul EC2, vous devez fournir des configurations d'encoches qui spécifient les [tailles d'instance EC2](#) que vous souhaitez que chaque serveur fournisse. AWS configure chaque serveur avec le schéma d'encodage demandé.

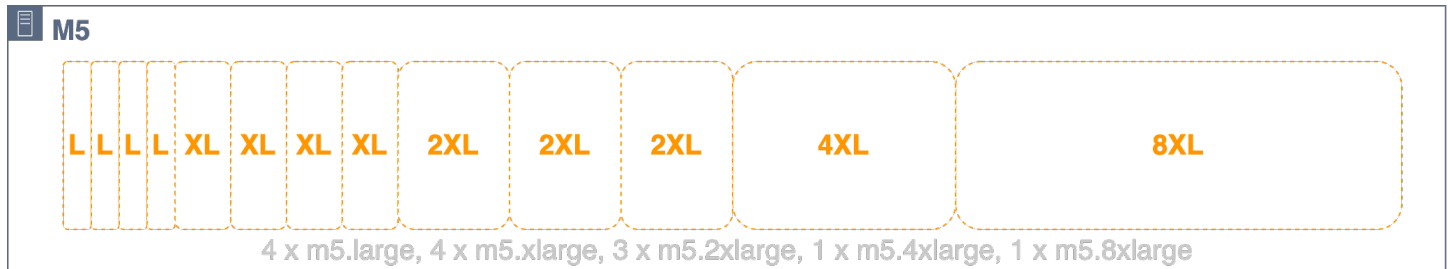
Les serveurs peuvent être répartis de manière homogène lorsque tous les emplacements ont la même taille d'instance (par exemple, 48 m5.large emplacements) ou répartis de manière hétérogène avec un mélange de types d'instances (par exemple, 4, 4m5.large, 3 m5.xlarge m5.2xlarge m5.4xlarge, 1 et 1m5.8xlarge). Consultez les trois figures suivantes pour des visualisations de ces configurations de créneaux.



## m5.24xlarge ressources informatiques du serveur



### m5.24xlarge serveur réparti de manière homogène sur 48 emplacements m5.large



### m5.24xlarge serveur réparti de manière hétérogène en 4m5.large, 4, 3 m5.xlarge

### m5.2xlarge m5.4xlarge, 1 et 1 emplacements m5.8xlarge

Il n'est pas nécessaire de répartir la capacité totale du serveur. Des emplacements peuvent être ajoutés à un serveur qui dispose d'une capacité non allouée disponible. Vous pouvez modifier un schéma de créneaux en ouvrant un ticket d'assistance. Le Support aux entreprises peut vous demander d'arrêter ou de redémarrer certaines instances pour terminer une demande de relocalisation si le nouveau schéma de créneaux ne peut pas être appliqué alors que certains emplacements sont occupés par des instances en cours d'exécution.

Tous les serveurs fournissent leurs emplacements provisionnés aux pools de capacité EC2 de l'Outpost, et tous les emplacements d'un type et d'une taille d'instance donnés sont gérés comme un pool de capacité EC2 unique. Par exemple, le précédent serveur à emplacements hétérogènes avec des emplacements m5.large, m5.xlarge m5.2xlarge m5.4xlarge, et distribuait ces m5.8xlarge emplacements à cinq pools de capacité EC2, soit un pool pour chaque type et taille d'instance.

Il est important de tenir compte de l'emplacement des serveurs et des pools de capacité EC2 lors de la planification de la capacité de réserve pour la disponibilité des serveurs N+M. AWS détecte la défaillance ou la dégradation d'un serveur et planifie une visite sur site pour remplacer le serveur défaillant. Vous devez concevoir vos pools de capacité EC2 de manière à tolérer la défaillance d'au

moins un serveur de chaque famille d'instances (N+1) dans un Outpost. Avec ce niveau minimum de disponibilité des serveurs, lorsqu'un serveur tombe en panne ou doit être mis hors service, vous pouvez redémarrer les instances défectueuses ou dégradées sur les emplacements de réserve des serveurs restants de la même famille.

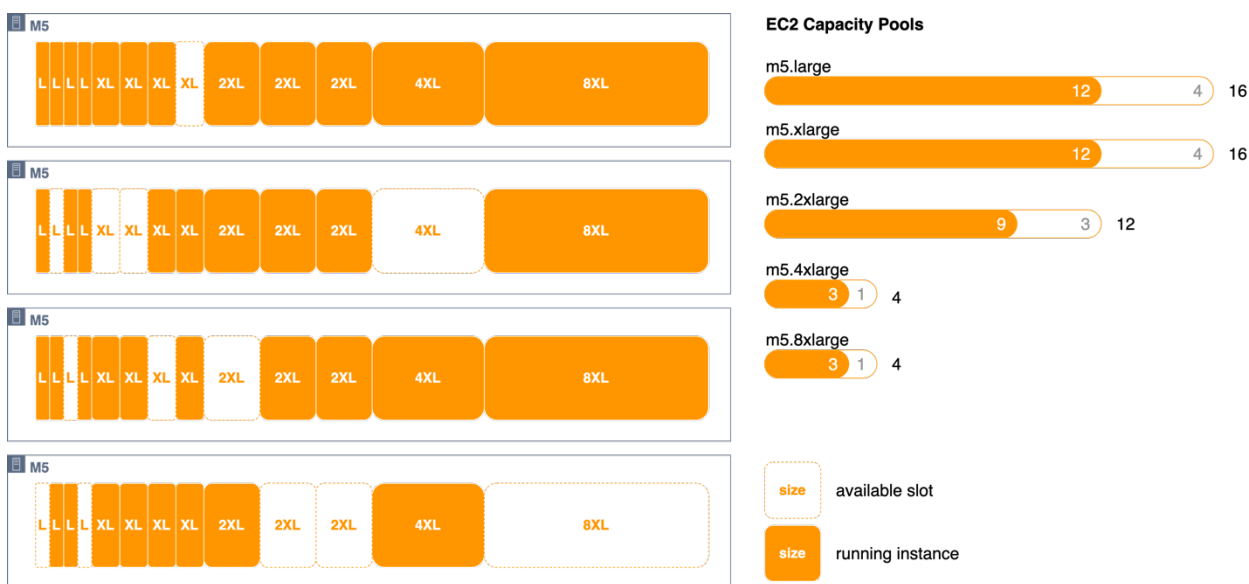
La planification de la disponibilité de N+M est simple lorsque vous disposez de serveurs à emplacements homogènes ou de groupes de serveurs à emplacements hétérogènes avec des configurations d'emplacements identiques. Il vous suffit de calculer le nombre de serveurs (N) dont vous avez besoin pour exécuter toutes vos charges de travail, puis d'ajouter (M) de serveurs supplémentaires pour répondre à vos exigences en matière de disponibilité des serveurs en cas de panne ou de maintenance.

Les configurations de créneaux suivantes ne sont pas utilisables en raison des limites NUMA :

- 3 m5.8xlarge
- 1 m5.16xlarge et 1 m5.8xlarge

Consultez votre Compte AWS équipe pour valider la configuration d'encodage des AWS Outposts racks prévue.

Dans la figure suivante, quatre m5.24xlarge serveurs sont répartis de manière hétérogène avec un schéma d'encodage identique. Les quatre serveurs créent cinq pools de capacité EC2. Chaque pool fonctionne au taux d'utilisation maximal (75 %) afin de maintenir une disponibilité N+1 pour les instances exécutées sur ces quatre serveurs. En cas de défaillance d'un serveur, il y a suffisamment d'espace pour redémarrer les instances défectueuses sur les serveurs restants.



## Visualisation des emplacements de serveur EC2, des instances en cours d'exécution et des pools d'emplacements

Pour les configurations de créneaux plus complexes, dans lesquelles les serveurs ne sont pas répartis de la même manière, vous devez calculer la disponibilité N+M pour chaque pool de capacités EC2. Vous pouvez utiliser la formule suivante pour calculer le nombre de serveurs (qui fournissent des emplacements à un pool de capacité EC2 donné) susceptibles de tomber en panne tout en permettant aux serveurs restants de transporter les instances en cours d'exécution :

$$M = \left\lceil \frac{\text{poolSlots}_{\text{available}}}{\text{serverSlots}_{\text{max}}} \right\rceil$$

Où :

- $\text{PoolSlots}_{\text{available}}$  est le nombre d'emplacements disponibles dans le pool de capacité EC2 donné (nombre total d'emplacements dans le pool moins le nombre d'instances en cours d'exécution)
- $\text{ServerSlots}_{\text{max}}$  est le nombre maximal d'emplacements fournis par un serveur au pool de capacités EC2 donné.
- M est le nombre de serveurs susceptibles de tomber en panne tout en permettant aux serveurs restants de transporter les instances en cours d'exécution

Exemple : un avant-poste possède trois serveurs qui fournissent des emplacements à un pool `m5.2xlarge` de capacités. Le premier fournit 4 emplacements, le second 3 emplacements et le troisième serveur 2 emplacements. Le pool d'`m5.2xlarge` instances de l'Outpost a une capacité totale de 9 emplacements (4 + 3 + 2). The Outpost dispose de 4 `m5.2xlarge` instances en cours d'exécution. Combien de serveurs risquent de tomber en panne tout en permettant aux serveurs restants de transporter les instances en cours d'exécution ?

$$poolSlots_{available} = total\ capacity - running\ instances = 9 - 4 = 5$$

$$serverSlots_{max} = \max([4, 3, 2]) = 4$$

$$M = \left\lceil \frac{poolSlots_{available}}{serverSlots_{max}} \right\rceil = \left\lceil \frac{5}{4} \right\rceil = [1.25] = 1$$

Réponse : Vous pouvez perdre n'importe lequel des serveurs tout en conservant les instances en cours d'exécution sur les serveurs restants.

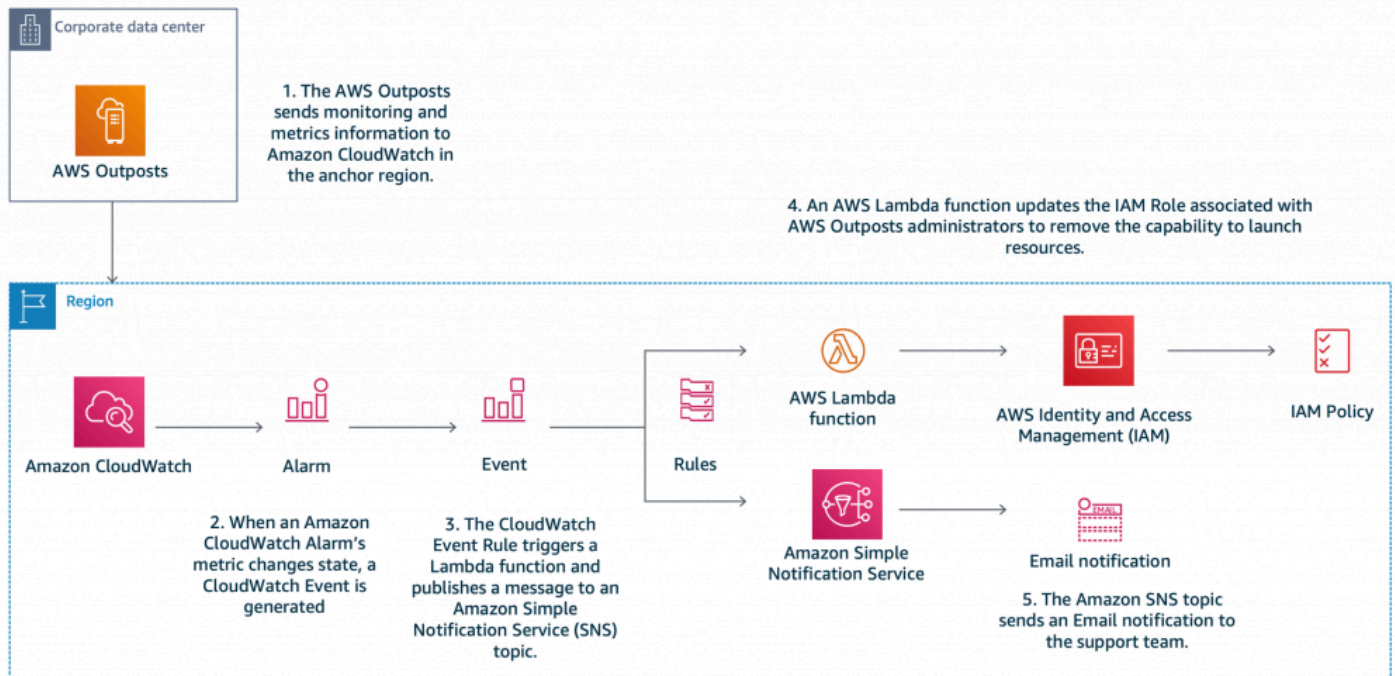
Pratiques recommandées pour la planification de la capacité de calcul :

- Dimensionnez votre capacité de calcul de manière à fournir une redondance N+M pour chaque pool de capacités EC2 d'un avant-poste.
- Déployez des serveurs N+M pour des serveurs homogènes ou identiques répartis de manière hétérogène.
- Calculez la disponibilité N+M pour chaque pool de capacités EC2 et assurez-vous que chaque pool répond à vos exigences de disponibilité.

## Gestion de capacité

Vous pouvez surveiller l'utilisation du pool d'instances Outpost EC2 dans AWS Management Console et via les métriques Amazon CloudWatch. Contactez le Support aux entreprises pour récupérer ou modifier les configurations de créneaux de vos Outposts.

Vous utilisez les mêmes mécanismes de [restauration automatique d'instance](#) et d'[EC2 Auto Scaling](#) pour récupérer ou remplacer les instances touchées par des pannes de serveur et des événements de maintenance. Vous devez surveiller et gérer la capacité de votre avant-poste afin de vous assurer qu'une capacité de réserve suffisante est toujours disponible pour faire face aux pannes de serveur. Le billet de AWS Lambda blog intitulé [Gérer votre AWS Outposts capacité à l'aide d'Amazon CloudWatch](#) propose un didacticiel pratique vous expliquant comment combiner AWS CloudWatch et AWS Lambda gérer les capacités de votre Outpost afin de garantir la disponibilité des instances.



## Gérer les AWS Outposts capacités avec Amazon CloudWatch et AWS Lambda

### Pratiques recommandées pour la gestion de la capacité de calcul :

- Configurez vos instances EC2 dans des groupes Auto Scaling ou utilisez la restauration automatique des instances pour redémarrer les instances défectueuses.
- Automatisez la surveillance de la capacité pour vos déploiements Outpost et configurez des notifications et (éventuellement) des réponses automatisées pour les alarmes de capacité.

## Placement de l'instance

Les Outposts disposent d'un nombre limité de serveurs informatiques. Si votre application déploie plusieurs instances associées sur Outposts ; sans configuration supplémentaire, les instances peuvent être déployées sur le même serveur ou sur des serveurs du même rack. Aujourd'hui, il existe trois mécanismes que vous pouvez utiliser pour distribuer des instances afin d'atténuer le risque lié à l'exécution d'instances associées sur la même infrastructure :

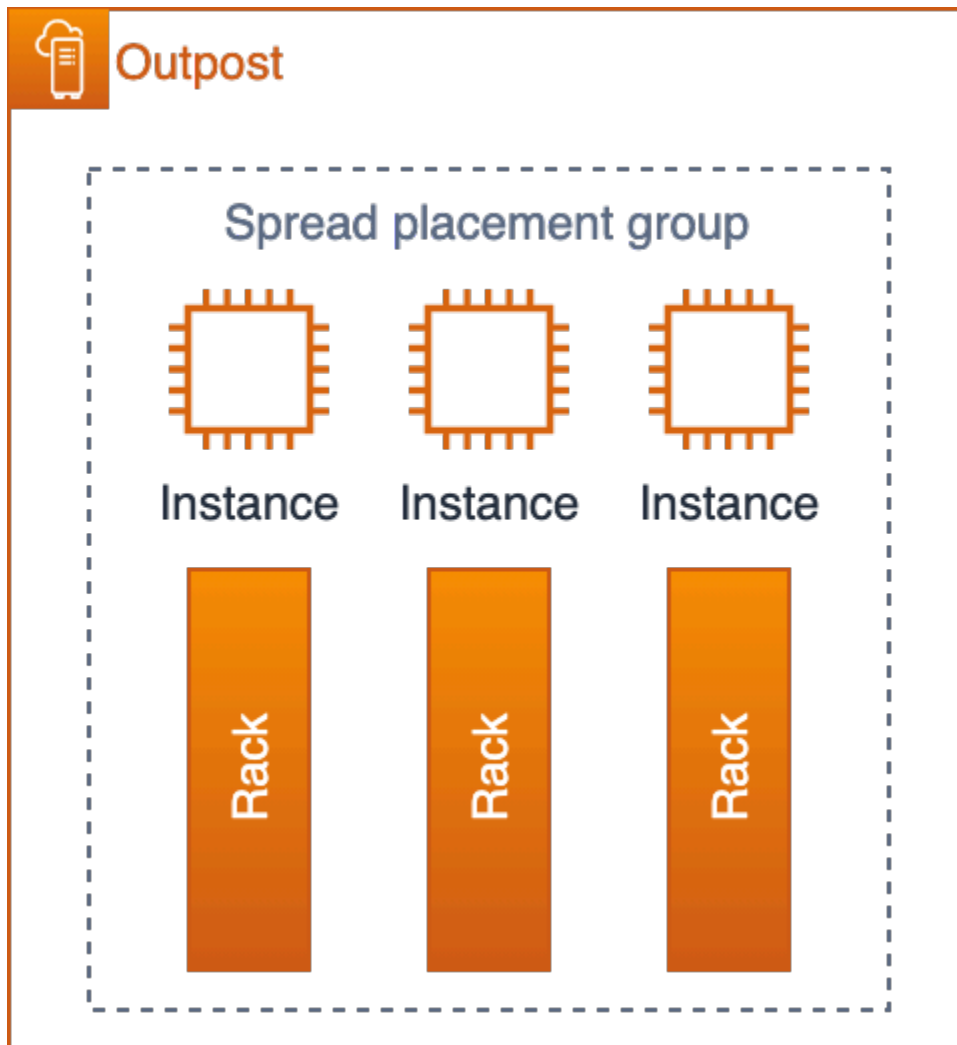
Déploiement de plusieurs Outposts : à l'instar d'une stratégie multi-AZ dans la région, vous pouvez déployer des Outposts dans des centres de données distincts et déployer des ressources applicatives sur des Outposts spécifiques. Cela vous permet d'exécuter des instances sur l'Outpost souhaité (un ensemble logique de racks). Une stratégie multi-avant-postes peut être utilisée pour se



protéger contre les modes de défaillance des racks et des centres de données et, si les Outposts sont ancrés dans des zones ou des régions distinctes, elle peut également fournir une protection contre les modes de défaillance de l'AZ ou de la région. Pour plus d'informations sur les architectures Multi-Outpost, consultez les Modes de [défaillance les plus importants](#).

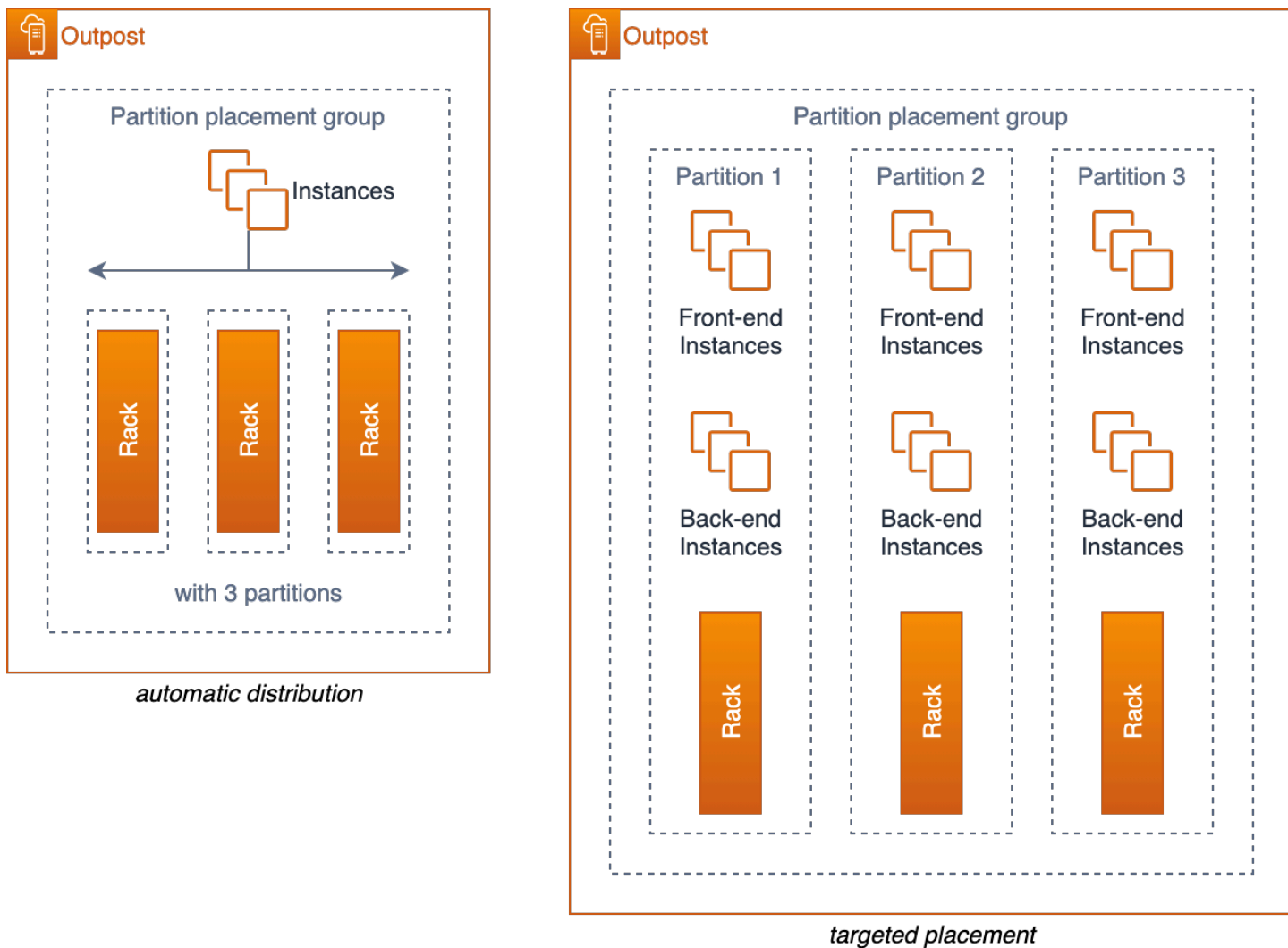
[Les groupes de placement Amazon EC2 sur les Outposts \(placement d'instances sur plusieurs racks à avant-poste unique\) vous permettent d'utiliser les stratégies de cluster, de propagation et de partition pour influencer le placement.](#) Les stratégies de répartition et de placement des partitions vous permettent de répartir les instances sur plusieurs racks dans un avant-poste à plusieurs racks.

Un groupe de placement de spread fournit un moyen simple de distribuer des instances uniques sur des racks afin de réduire le risque de défaillances corrélées. Vous ne pouvez déployer dans le groupe qu'autant d'instances que vous avez de racks dans votre avant-poste.



Groupe de placement de spread EC2 sur un avant-poste avec trois racks

Vous pouvez également répartir les instances sur plusieurs racks avec des groupes de placement de partitions. Utilisez la distribution automatique pour répartir les instances sur les partitions du groupe ou pour déployer des instances sur des partitions cibles sélectionnées. Le déploiement d'instances sur des partitions cibles vous permet de déployer des ressources sélectionnées sur le même rack tout en répartissant les autres ressources entre les racks. Par exemple, si vous avez un avant-poste logique doté de trois racks, la création d'un groupe de placement de partitions avec trois partitions vous permet de répartir les ressources entre les racks.



### Groupes de placement de partitions EC2 sur un avant-poste doté de trois racks

Emplacement des serveurs Creative : si vous avez un Outpost à rack unique ou si le service que vous utilisez sur Outposts ne prend pas en charge les groupes de placement, vous pouvez peut-être utiliser le slot créatif pour vous assurer que vos instances ne sont pas déployées sur le même serveur physique. Si les instances associées ont la même taille d'instance EC2, vous pouvez peut-être attribuer des emplacements à vos serveurs afin de limiter le nombre d'emplacements de cette

taille configurés sur chaque serveur, en répartissant les emplacements sur les serveurs. Le slotting du serveur limitera le nombre d'instances (de cette taille) pouvant être exécutées sur un seul serveur.

À titre d'exemple, considérez le schéma de rainures illustré précédemment dans la Figure 13. Si votre application devait déployer trois `m5.4xlarge` instances sur l'Outpost configuré avec ce schéma de créneaux, EC2 placerait chaque instance sur un serveur distinct et il n'y aurait aucune possibilité que ces instances puissent fonctionner sur le même serveur, à condition que la configuration des créneaux ne change pas pour ouvrir des `m5.4xlarge` emplacements supplémentaires sur les serveurs.

Pratiques recommandées pour le placement des instances de calcul :

- Utilisez les groupes de placement Amazon EC2 sur les Outposts pour contrôler le placement des instances entre les racks au sein d'un même Outpost.
- Au lieu de commander un Outpost avec un seul rack Outpost de taille moyenne ou grande, envisagez de diviser la capacité en deux racks de petite ou moyenne taille afin de tirer parti de la capacité des groupes de placement EC2 à répartir les instances sur les racks.

## Stockage

Le service de AWS Outposts rack propose trois types de stockage :

- [Stockage d'instance](#) sur les types d'instances EC2 pris en charge
- [Volumes Amazon Elastic Block Store \(EBS\) gp2](#) pour le stockage par blocs persistant
- [Amazon Simple Storage Service on Outposts \(S3 on Outposts\)](#) pour le stockage d'objets local

Le stockage des instances est fourni sur les serveurs pris en charge (`C5dM5dR5d`, `G4dn`, et `I3en`). Tout comme dans la région, les données d'un magasin d'instance ne sont conservées que pendant la [durée de vie \(courante\) de l'instance](#).

Les volumes EBS d'Outposts et le stockage d'objets S3 on Outposts sont fournis dans le cadre des services gérés en rack. AWS Outposts Les clients sont responsables de la gestion de la capacité des pools de stockage Outpost. Les clients spécifient leurs besoins de stockage pour le stockage EBS et S3 lorsqu'ils commandent un Outpost. AWS configure l'Outpost avec le nombre de serveurs de stockage requis pour fournir la capacité de stockage demandée. AWS est responsable de la disponibilité des services de stockage EBS et S3 on Outposts. Un nombre suffisant de serveurs de stockage est configuré pour fournir des services de stockage hautement disponibles à l'Outpost.

La perte d'un seul serveur de stockage ne doit pas perturber les services ni entraîner de perte de données.

Vous pouvez utiliser les [CloudWatch métriques AWS Management Console](#) et pour surveiller l'utilisation des capacités d'Outpost EBS et [S3 on Outposts](#).

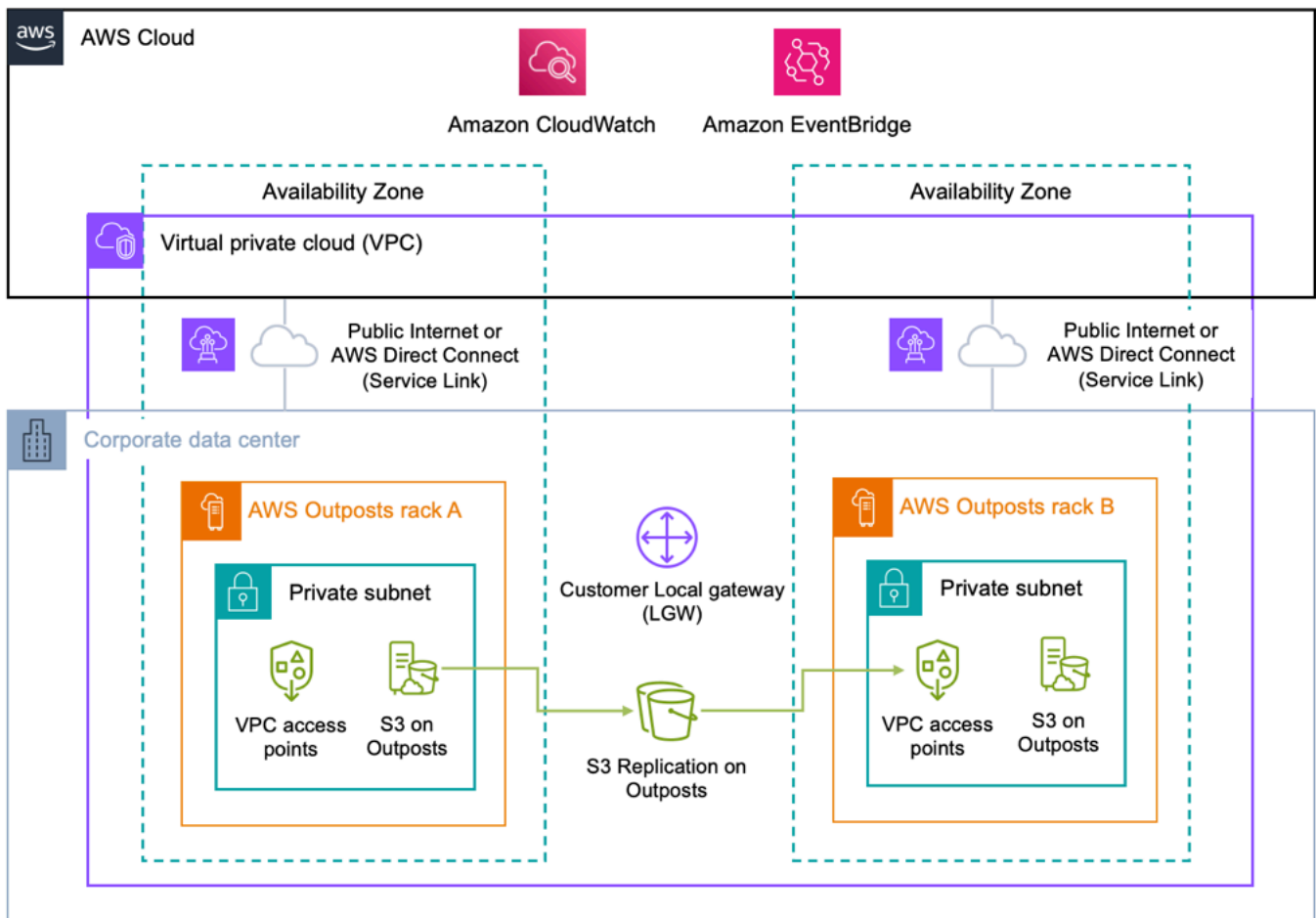
## Protection des données

Pour les volumes EBS : le AWS Outposts rack prend en charge les instantanés de volumes EBS afin de fournir un mécanisme de protection des données simple et sécurisé afin de protéger vos données de stockage par blocs. Les snapshots sont des sauvegardes point-in-time incrémentielles de vos volumes EBS. Par défaut, les [instantanés des volumes Amazon EBS](#) sur votre Outpost sont stockés sur Amazon S3 dans la région. Si vos Outposts ont été configurés avec la capacité S3 on Outposts, vous pouvez utiliser [EBS Local Snapshots on Outposts pour stocker des instantanés localement sur votre Outpost à l'aide du stockage S3 on Outposts](#).

Pour les compartiments S3 on Outposts (cas d'utilisation de la résidence des données) :

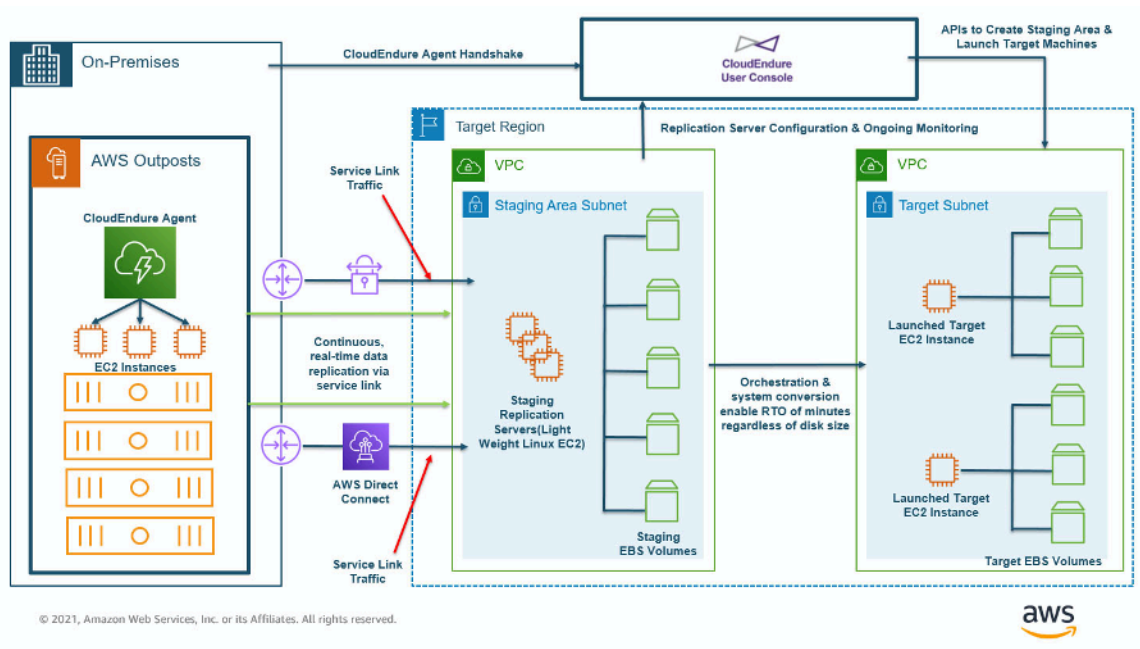
- Vous pouvez utiliser le contrôle de [version S3 sur les](#) Outposts pour enregistrer toutes les modifications et l'historique des objets. Une fois activé, la gestion des versions S3 enregistre plusieurs copies différentes d'un objet dans le même compartiment. Vous pouvez utiliser la gestion des versions S3 pour préserver, récupérer et restaurer chaque version de chaque objet stocké dans votre compartiment Outposts. La gestion des versions S3 vous aide à récupérer en cas d'action involontaire d'un utilisateur et de défaillance applicative.
- Vous pouvez utiliser la [réplication S3 sur les Outposts](#) pour créer et configurer des règles de réplication afin de répliquer automatiquement vos objets S3 vers un autre Outpost ou vers un autre bucket du même Outpost. Pendant la réplication, les objets S3 on Outposts sont envoyés via la passerelle locale (LGW) du client, et les objets ne retournent pas vers le. Région AWS S3 Replication on Outposts fournit un moyen simple et flexible de répliquer automatiquement les données au sein d'un périmètre de données spécifique afin de répondre aux exigences de redondance et de conformité des données.

S3 Replication on Outposts fournit également des métriques et des notifications détaillées pour surveiller l'état de la réplication de vos objets. Vous pouvez suivre la progression de la réplication en suivant les octets en attente, les opérations en attente et la latence de réplication entre vos compartiments Outposts source et de destination à l'aide d'Amazon CloudWatch. Vous pouvez également configurer des EventBridge règles Amazon pour recevoir les événements d'échec de réplication afin de diagnostiquer et de corriger rapidement les problèmes de configuration.



Pour les compartiments S3 on Outposts (cas d'utilisation autres que la résidence de données) à Régions AWS : Vous pouvez les utiliser [AWS DataSync](#) pour automatiser les transferts de données S3 on Outposts entre votre Outpost et la région. DataSync vous permet de choisir les éléments à transférer, le moment du transfert et la quantité de bande passante à utiliser. La sauvegarde de vos compartiments S3 on Outposts sur site vers des compartiments S3 vous permet de tirer parti de Région AWS la durabilité des données de 99,999999999 % (11 9) et des niveaux de stockage supplémentaires (Standard, Infrequent Access et Glacier) pour optimiser les coûts grâce au service S3 régional.

Réplication d'instances : vous pouvez l'utiliser [CloudEndure](#) pour répliquer des instances individuelles depuis des systèmes sur site vers un avant-poste, d'un avant-poste vers la région, de la région vers un avant-poste ou d'un avant-poste à un autre. Le billet de CloudEndure blog [Architecting for DR on AWS Outposts with](#) décrit chacun de ces scénarios et explique comment concevoir une solution avec CloudEndure.



## Reprise après sinistre (DR) depuis un avant-poste vers la région

L'utilisation du AWS Outposts rack comme CloudEndure destination (cible de réplication) nécessite le stockage S3 on Outposts.

## Pratiques recommandées pour la protection des données :

- Utilisez les instantanés EBS pour créer point-in-time des sauvegardes de volumes de stockage par blocs vers Amazon S3 dans la région ou S3 sur Outposts.
- Utilisez le versionnement des objets S3 on Outposts pour conserver plusieurs versions et l'historique de vos objets.
- Utilisez S3 Replication on Outposts pour répliquer automatiquement les données de vos objets vers un autre Outpost.
- Pour les cas d'utilisation autres que la résidence de données, utilisez-le AWS DataSync pour sauvegarder les objets stockés dans S3 sur Outpost sur Amazon S3 dans la région.
- CloudEndure À utiliser pour répliquer des instances entre des systèmes sur site, des Outposts logiques et la région.

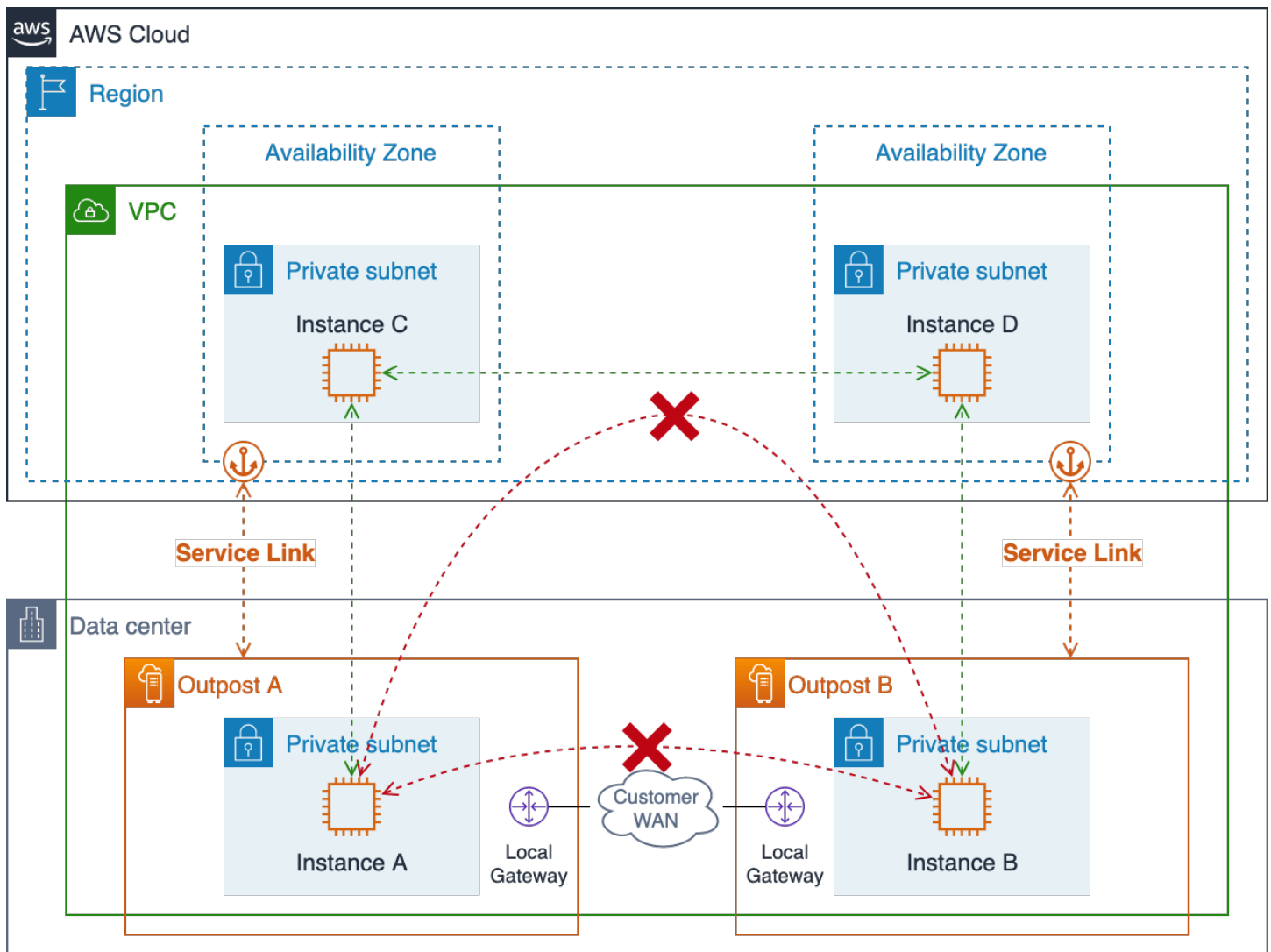
## Modes de défaillance plus importants

Pour concevoir des architectures HA afin d'atténuer les défaillances les plus importantes, telles que les pannes de rack, de centre de données, de zone de disponibilité (AZ) ou de région, vous devez

déployer plusieurs Outposts dotés d'une capacité d'infrastructure suffisante dans des centres de données distincts dotés d'une alimentation et d'une connectivité WAN indépendantes. Vous ancrez les Outposts dans différentes zones de disponibilité (AZ) au sein d'une Région AWS ou de plusieurs régions. Vous devez également fournir une site-to-site connectivité résiliente et suffisante entre les sites pour prendre en charge la réplication synchrone ou asynchrone des données et la redirection du trafic de charge de travail. En fonction de l'architecture de votre application, vous pouvez utiliser le DNS [Amazon Route 53](#) disponibles dans le monde entier et les services [Elastic Load Balancing](#) disponibles au niveau régional pour diriger le trafic vers l'emplacement souhaité et automatiser la redirection du trafic vers les sites survivants en cas de panne à grande échelle.

Il existe des limites du réseau dont vous devez tenir compte lorsque vous concevez et déployez des charges de travail d'applications sur plusieurs Outposts. Les ressources de deux Outposts distincts ne peuvent pas communiquer entre elles en faisant transiter le trafic par la Région. Les ressources de deux Outposts distincts déployés au sein d'un même VPC ne peuvent pas communiquer entre elles sur le réseau du client. Les ressources de deux Outposts distincts déployés dans des VPC différents peuvent communiquer entre elles via le réseau du client.

Les deux figures suivantes illustrent les chemins réseau bloqués et réussis.



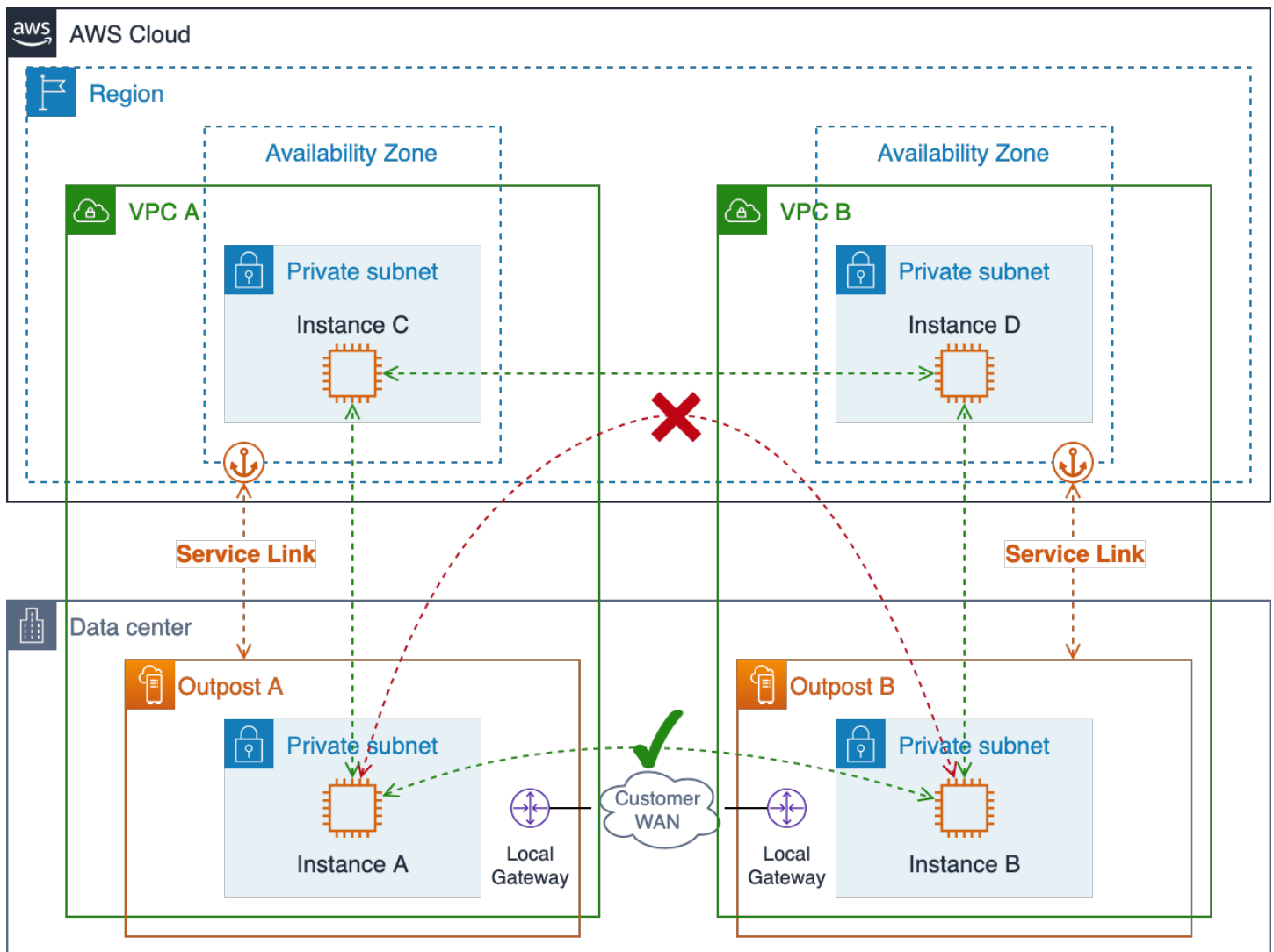
### Chemins réseau à plusieurs avant-postes VPC uniques

Le trafic d'avant-poste à avant-poste transitant par la Région est bloqué car il s'agit d'un antischéma. Un tel trafic entraînerait des frais de sortie dans les deux sens et aurait probablement une latence beaucoup plus élevée que le simple routage du trafic sur le réseau étendu du client.

Les ressources de plusieurs Outposts d'un même VPC ne peuvent pas communiquer entre elles. Le trafic entre Outpost dans le même VPC suivra toujours la route CIDR VPC locale à travers la région où il sera bloqué.

Vous devez utiliser des VPC distincts pour déployer des ressources sur plusieurs Outposts afin de vous permettre d'acheminer le trafic d'Outpost à Outpost sur vos réseaux locaux sur site et WAN.





### Chemins réseau à plusieurs VPC et à plusieurs avant-postes

Pratiques recommandées pour se protéger contre les modes de défaillance plus importants :

- Déployez plusieurs Outposts ancrés dans plusieurs AZ et régions.
- Utilisez des VPC distincts pour chaque avant-poste dans le cadre d'un déploiement multi-avant-poste.

## Conclusion

Avec le AWS Outposts rack, vous pouvez créer, gérer et faire évoluer des applications sur site hautement disponibles à l'aide d' AWS outils et de services courants tels qu'Amazon EC2, Amazon EBS, Amazon S3 on Outposts, Amazon ECS, Amazon EKS et Amazon RDS. Les charges de travail peuvent être exécutées localement, servir les clients, accéder aux applications et aux systèmes de vos réseaux locaux et accéder à l'ensemble complet des services du. Région AWS Le rack Outposts est idéal pour les charges de travail qui nécessitent un accès à faible latence aux systèmes sur site, le traitement local des données, la résidence des données et la migration d'applications avec des interdépendances entre les systèmes locaux.

Lorsque vous fournissez à un déploiement Outpost une alimentation, un espace et un refroidissement adéquats et des connexions résilientes au Région AWS, vous pouvez créer des services de centre de données uniques hautement disponibles. Et pour des niveaux supérieurs de disponibilité et de résilience, vous pouvez déployer plusieurs Outposts et distribuer vos applications au-delà des limites logiques et géographiques.

Outposts Rack élimine le fardeau indifférencié lié à la création de pools de calcul, de stockage et de réseaux d'applications sur site et vous permet d'étendre la portée de l'infrastructure AWS mondiale à vos centres de données et à vos installations de colocation. Vous pouvez désormais consacrer votre temps et votre énergie à la modernisation de vos applications, à la rationalisation de vos déploiements d'applications et à l'augmentation de l'impact commercial de vos services informatiques.

# Collaborateurs

Les personnes qui ont contribué à ce document incluent :

- Mallory Gershenfeld, S3 sur les Outposts, Amazon Web Services
- Chris Lunsford, architecte de solutions spécialisé senior AWS Outposts, Amazon Web Services
- Rohan Mathews, architecte en chef AWS Outposts, Amazon Web Services

# Historique du document

Pour être informé des mises à jour de ce livre blanc, abonnez-vous au flux RSS.

Modification	Description	Date
<a href="#">Mise à jour mineure</a>	Ajout de conseils de créneaux supplémentaires dans le cadre de la planification des capacités.	9 février 2024
<a href="#">Mise à jour mineure</a>	Mis à jour pour refléter les lancements de fonctionnalités depuis la publication initiale.	19 juillet 2023
<a href="#">Mise à jour mineure</a>	Pratiques recommandées mises à jour pour la connexion réseau à haute disponibilité.	29 juin 2023
<a href="#">Publication initiale</a>	Livre blanc publié pour la première fois.	12 août 2021

## Note

Pour vous abonner aux mises à jour RSS, un plug-in RSS doit être activé pour le navigateur que vous utilisez.

## Avis

Il incombe aux clients de procéder à une évaluation indépendante des informations contenues dans le présent document. Ce document : (a) est fourni à titre informatif uniquement, (b) représente les offres de AWS produits et les pratiques actuelles, qui sont susceptibles d'être modifiées sans préavis, et (c) ne crée aucun engagement ni aucune assurance de la part de AWS ses filiales, fournisseurs ou concédants de licence. AWS les produits ou services sont fournis « tels quels » sans garanties, déclarations ou conditions d'aucune sorte, qu'elles soient explicites ou implicites. Les responsabilités et obligations AWS de ses clients sont régies par AWS des accords, et ce document ne fait partie d'aucun accord conclu entre AWS et ses clients et ne les modifie pas.

© 2023 Amazon Web Services, Inc. ou ses filiales. Tous droits réservés.

# AWS Glossaire

Pour la AWS terminologie la plus récente, consultez le [AWS glossaire](#) dans la Glossaire AWS référence.

Les traductions sont fournies par des outils de traduction automatique. En cas de conflit entre le contenu d'une traduction et celui de la version originale en anglais, la version anglaise prévaudra.