Approches de sauvegarde et de récupération avec AWS

Juin 2016

Mentions légales

Ce document est fourni à titre informatif uniquement. Il présente l'offre de produits et les pratiques actuelles d'AWS à la date de publication de ce document, des informations qui sont susceptibles d'être modifiées sans avis préalable. Il incombe aux clients de procéder à leur propre évaluation indépendante des informations contenues dans ce document et chaque client est responsable de son utilisation des produits ou services AWS, chacun étant fourni « en l'état », sans garantie d'aucune sorte, qu'elle soit explicite ou implicite. Ce document ne crée pas de garanties, représentations, engagements contractuels, conditions ou assurances à l'encontre d'AWS, de ses affiliés, fournisseurs ou donneurs de licence. Les responsabilités et obligations d'AWS vis-à-vis de ses clients sont régies par les contrats AWS. Le présent document ne fait partie d'aucun contrat et ne modifie aucun contrat entre AWS et ses clients.

Table des matières

Résumé

Ce document s'adresse aux architectes solutions, architectes sauvegardes et administrateurs informatiques de l'entreprise qui sont en charge de la protection des données pour les environnements informatiques de leur société. Il détaille les charges de travail de production et les architectures pouvant être implémentées à l'aide d'AWS pour renforcer ou remplacer une solution de sauvegarde et de récupération. Ces approches proposent des coûts moins élevés, une évolutivité plus importante et plus de durabilité pour satisfaire aux exigences en termes d'objectif de délai de récupération (RTO), d'objectif de point de récupération (RPO) et de conformité.

Introduction

Alors que la croissance des données d'entreprise accélère, leur protection devient plus difficile. Les questions de durabilité et d'évolutivité des méthodes de sauvegarde font désormais partie du quotidien, notamment celle-ci : Comment le cloud permet-il de répondre à mes besoins de sauvegarde et d'archivage ?

Cet article couvre plusieurs architectures de sauvegarde (applications résidant dans le cloud, environnements hybrides et sur site) et les services AWS associés qui peuvent être utilisés pour créer des solutions de protection de données évolutives et fiables.

Pourquoi utiliser AWS comme plateforme de protection des données ?

Amazon Web Services (AWS) est une plateforme de cloud computing flexible, économique, sécurisée, à haute performance et simple à utiliser. AWS se charge des lourdes tâches indifférenciées et fournit les outils et les ressources que vous pouvez utiliser pour créer des solutions évolutives de sauvegarde et de récupération.

L'utilisation d'AWS dans le cadre de votre stratégie de protection des données présente de nombreux avantages :

- **Durabilité :** Amazon Simple Storage Service (Amazon S3) et Amazon Glacier sont conçus pour être à 99,999999999 % (lisez « onze-neuf ») durables pour les objets qu'ils stockent. Les deux plateformes proposent des emplacements fiables pour la sauvegarde de données.

- **Sécurité :** AWS fournit plusieurs options de contrôle d'accès et de chiffrement des données en transit et au repos.
- **Infrastructure globale :** Les services AWS sont disponibles à travers le monde de sorte que vous puissiez sauvegarder et stocker vos données dans la région qui répond à vos impératifs de conformité.
- **Conformité :** L'infrastructure AWS infrastructure est conforme aux normes telles que Service Organization Controls (SOC), Statement on Standards for Attestation Engagements (SSAE) No. 16, Organisation internationale de normalisation (ISO) 27001, la norme de sécurité dans le secteur des cartes de paiement (PCI DSS), Health Insurance Portability and Accountability Act (HIPPA), SEC[1] et Federal Risk and Authorization Management Program (FedRAMP) pour que vous puissiez intégrer facilement la solution de sauvegarde à votre régime de conformité existant.
- **Évolutivité :** Avec AWS, la capacité n'est plus un souci. Vous pouvez augmenter ou diminuer votre consommation selon l'évolution de vos besoins sans frais administratifs.
- **Diminution du coût total de possession :** La mise à l'échelle des opérations d'AWS réduit les coûts de service et favorise la diminution du coût total de possession du stockage. AWS répercute ces économies de coûts sur les clients sous forme de baisses de prix.
- **Tarification à l'utilisation :** Achetez des services AWS lorsque vous en avez besoin et uniquement pour la période pendant laquelle vous pensez les utiliser. Les tarifs AWS n'exigent pas de frais initiaux, de résiliation ou de contrats à long terme.

Services de stockage AWS pour la protection des données

Amazon S3 et Amazon Glacier sont des services idéaux pour la sauvegarde et l'archivage. Il s'agit de plateformes de stockage durables et économiques. Elles proposent toutes les deux une capacité illimitée et ne nécessitent aucune gestion du volume ou multimédia lorsque les ensembles de données de sauvegarde se développent. Le modèle de paiement uniquement à l'utilisation et le faible coût par Go/mois rendent ces services parfaitement adaptés aux cas d'utilisation de protection des données.

[1] https://aws.amazon.com/about-aws/whats-new/2015/09/amazon-glacier-receives-third-party-compliance-assessment-for-sec-rule-17a-4f-from-cohasset-associates-inc/

Amazon S3

Amazon S3 fournit un stockage d'objet extrêmement évolutif et sécurisé.

Vous pouvez utiliser Amazon S3 pour stocker et récupérer n'importe quel volume de données, à tout moment et depuis n'importe quel accès Internet. Amazon S3 stocke les données en tant qu'objets dans des ressources appelées *compartiments*. AWS Storage Gateway et plusieurs solutions de sauvegarde tierces peuvent gérer des objets Amazon S3 en votre nom. Vous pouvez stocker autant d'objets que vous souhaitez dans un compartiment et vous pouvez écrire, lire et supprimer les objets dans votre compartiment. La taille des objets uniques peut aller jusqu'à 5 To.

Amazon S3 offre toute une gamme de classes de stockage conçues pour différents cas d'utilisation. parmi lesquelles :

- **Amazon S3 Standard** pour un stockage à usage général des données fréquemment consultées.
- **Amazon S3 Standard-Infrequent Access** pour les données à longue durée de vie, mais moins fréquemment consultées.
- **Amazon Glacier** pour l'archivage à long terme.

Amazon S3 propose également des stratégies de cycle de vie que vous pouvez configurer pour gérer vos données tout au long de leur cycle de vie. Après la définition d'une stratégie, vos données seront migrées automatiquement vers la classe de stockage appropriée, sans aucune modification de votre application. Pour plus d'informations, consulte les classes de stockage S3.

Amazon Glacier

Amazon Glacier est un service très économique de stockage d'archives dans le cloud qui fournit un archivage et une sauvegarde en ligne des données sécurisés et durables. Pour conserver des prix bas, Amazon Glacier est optimisé pour les données auxquelles vous accédez de manière occasionnelle et pour lesquelles des délais d'extraction de plusieurs heures restent acceptables. Sans Amazon Glacier, vous pouvez stocker de façon fiable leurs volumes de données petits ou grands, pour seulement 0,007 USD par Go par mois, ce qui représente une économie importante comparée aux solutions sur site. Amazon Glacier est adapté pour le stockage de données de sauvegarde sans exigences relatives à la conservation longue ou indéfinie mais aussi pour l'archivage des données sur le long terme. Pour en savoir plus, consultez Amazon Glacier.

AWS Storage Gateway

AWS Storage Gateway connecte une application logicielle sur site à une unité de stockage basée sur le cloud afin de fournir une intégration continue et très sécurisée entre l'environnement informatique sur site et l'infrastructure de stockage AWS. Pour plus d'informations, consultez AWS Storage Gateway.

Services de transfert AWS

Outre les passerelles et les connecteurs tiers, vous pouvez utiliser les options AWS comme AWS Direct Connect, AWS Snowball, AWS Storage Gateway et l'accélération de transfert Amazon S3 afin de transférer rapidement vos données. Pour plus d'informations, consultez Cloud Data Migration.

Conception d'une solution de sauvegarde et de restauration

Lorsque vous développez une stratégie complète pour la sauvegarde et la restauration des données, vous devez tout d'abord identifier les situations d'échec et de sinistre qui peuvent survenir et leur impact commercial potentiel. Dans certains secteurs, vous devez prendre en compte les exigences réglementaires concernant la sécurité et la confidentialité des données ainsi que la conservation des enregistrements.

Vous devez implémenter les processus de sauvegarde qui offriront le niveau approprié de granularité pour répondre aux exigences de l'entreprise en matière d'objectif de durée de récupération (RTO) et d'objectif de point de récupération (RPO), notamment :

- Restauration au niveau du fichier
- Restauration au niveau du volume
- Restauration au niveau de l'application (par exemple, les bases de données)
- Restauration au niveau de l'image

Les sections suivantes décrivent les approches de sauvegarde, de récupération et d'archivage en fonction de l'organisation de votre infrastructure. L'infrastructure informatique peut se définir comme résidant dans le Cloud, sur site et hybride.

Infrastructure résidant dans le Cloud

Ce scénario décrit un environnement de charge de travail qui se trouve entièrement sur AWS. Comme la figure suivante le montre, il inclut les serveurs Web, les serveurs d'application, les serveurs de supervision, les bases de données et Active Directory.

Si vous exécutez tous vos services à partir d'AWS, vous pouvez tirer parti de plusieurs fonctions intégrées pour répondre à vos besoins en matière de protection et de récupération des données.

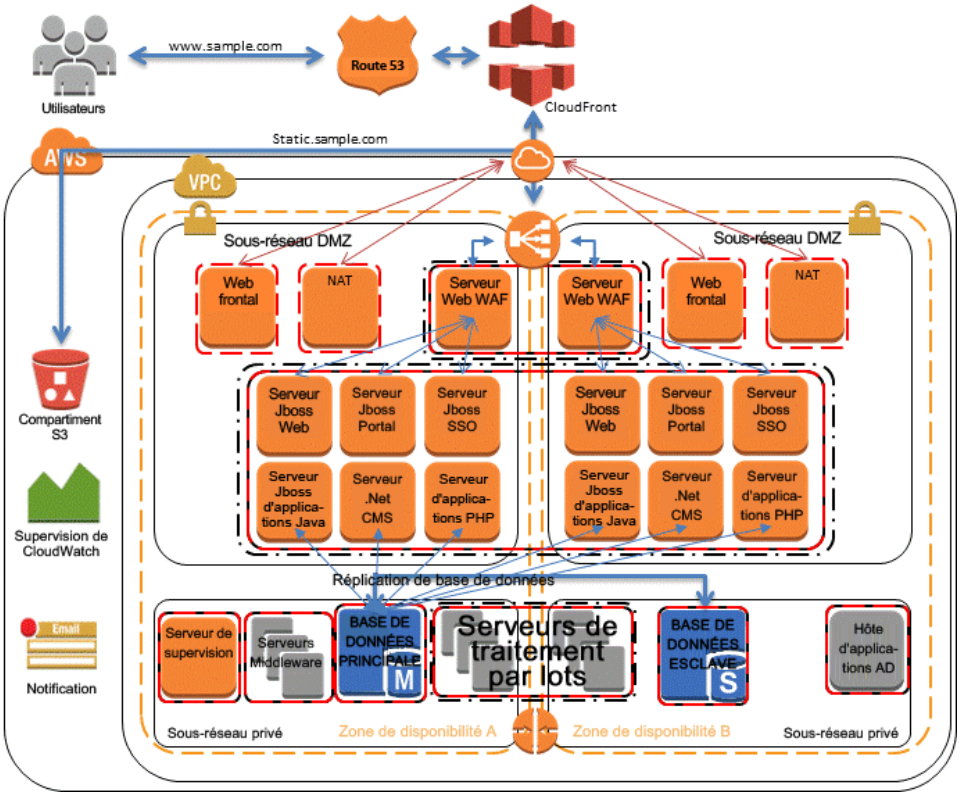


Figure1 : Scénario Résidant dans AWS

Protection basée sur l'instantané EBS

Lorsque les services sont exécutés dans Amazon Elastic Compute Cloud[2] (Amazon EC2), les instances de calcul peuvent utiliser les volumes Amazon Elastic Block Store (Amazon EBS) pour stocker et accéder aux données primaires. Vous pouvez utiliser ce stockage de bloc pour les données structurées, comme les bases de données, ou les données non structurées, comme les fichiers dans un système de fichiers sur le volume.

Amazon EBS permet de créer des instantanés (sauvegardes) de n'importe quel volume Amazon EBS. Il prend une copie du volume et la place dans Amazon S3 où elle est stockée de façon redondante dans plusieurs zones de disponibilité. Le premier instantané est une copie complète du volume. Les instantanés continus stockent uniquement les changements incrémentiels au niveau bloc.

Ce processus est un moyen rapide et fiable de restaurer les données de l'intégralité du volume. Pour une restauration partielle, il suffit d'attacher le volume à l'instance en cours d'exécution sous un nom de périphérique différent, le monter, puis utiliser les commandes du système d'exploitation pour copier les données depuis le volume de production.

Il est également possible de copier ces instantanés Amazon EBS entre les régions AWS qui utilisent la fonctionnalité de copie d'instantané d'Amazon EBS, disponible dans la console ou depuis la ligne de commande, tel que décrit dans le Guide de l'utilisateur d'Amazon Elastic Compute Cloud.[3] Vous pouvez utiliser cette fonction pour stocker votre sauvegarde dans une autre région sans avoir à gérer la technologie de réplication sous-jacente.

---

[2] http://aws.amazon.com/ec2/

[3] http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-copy-snapshot.html

Création des instantanés EBS

Quand vous créez un instantané, vous protégez vos données directement dans un stockage durable sur disque. Vous pouvez utiliser AWS Management Console, l'interface ligne de commande (CLI) ou les API pour créer l'instantané Amazon EBS.

Dans la console Amazon EC2, à la page **Elastic Block Store Volumes**, sélectionnez **Create Snapshot** dans le menu **Actions**. Dans la boîte de dialogue **Create Snapshot**, sélectionnez **Create** pour créer un instantané qui sera stocké dans Amazon S3.

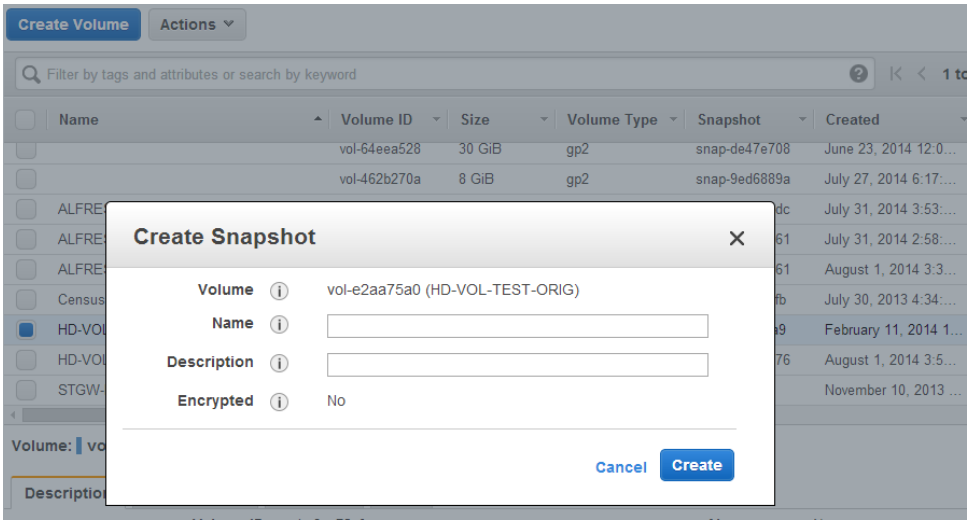


Figure2 : Utilisation de la console EC2 pour créer un instantané

Pour utiliser la commande CLI afin de créer un instantané, exécutez la commande suivante :

```
➢ aws ec2 create-snapshot
```

Vous pouvez planifier et exécuter les commandes `aws ec2 create-snapshot` de façon régulière pour sauvegarder les données EBS. La tarification économique d'Amazon S3 vous permet de conserver plusieurs générations de données. De plus, les instantanés étant basés sur des blocs, vous consommez uniquement l'espace correspondant aux données qui ont été modifiées par rapport à l'instantané initialement créé.

Restauration à partir d'un instantané EBS

Pour restaurer les données à partir d'un instantané, vous pouvez utiliser AWS Management Console, l'interface de ligne de commande (CLI) ou les API pour créer un volume depuis un instantané existant.

Par exemple, procédez comme suit pour restaurer un volume à partir d'une sauvegarde à un instant dans le passé :

1. Utilisez la commande suivante pour créer un volume à partir de l'instantané de sauvegarde :

```
➢ aws ec2 create-volume --region us-west-1b --snapshot-
  id mysnapshot-id
```

2. Sur l'instance Amazon EC2, démontez le volume existant.

   Sous Linux, utilisez `umount`. Sous Windows, utilisez le gestionnaire de volume logique (LVM).

3. Utilisez la commande suivante pour détacher le volume existant de l'instance :

```
➢ aws ec2 detach-volume --volume-id oldvolume-id –
  instance-id myec2instance-id
```

4. Utilisez la commande suivante pour attacher le volume qui a été créé à partir de l'instantané :

```
➢ aws ec2 attach-volume --volume-id newvolume-id
  --instance-id myec2instance-id --device /dev/sdf
```

5. Remontez le volume sur l'instance en cours d'exécution.

Création de sauvegardes cohérentes ou critiques

Quand vous effectuez une sauvegarde, l'idéal est que le système soit dans un mode où aucune opération d'E/S n'est effectuée. Idéalement, la machine n'accepte pas le trafic, mais cette situation est de plus en plus rare, la norme voulant aujourd'hui que les opérations informatiques s'effectuent 24h/24 et 7j/7.

Pour cette raison, vous devez mettre le système de fichiers ou la base de données au repos afin d'obtenir une sauvegarde propre. La méthode que vous utilisez dépend de votre base de données et/ou de votre système de fichiers.

Le processus pour une base de données se présente comme suit :

- Si possible, placez la base de données en mode de sauvegarde de secours.
- Exécutez les commandes de l'instantané Amazon EBS.
- Sortez la base de données du mode de sauvegarde de secours ou, si vous utilisez un réplica en lecture, résiliez l'instance réplica en lecture.

Le processus pour un système de fichiers est similaire, mais il dépend des capacités propres au système de fichiers ou au système d'exploitation. Par exemple, XFS est un système de fichiers pouvant vider ces données pour une sauvegarde cohérente. Pour plus d'informations, consultez xfs_freeze.[4]

Si votre système de fichiers n'a pas la capacité de geler les données, nous vous recommandons de le démonter, de générer la commande d'instantané, puis de remonter le système de fichiers. Vous pouvez également simplifier ce processus en utilisant le gestionnaire de volumes logiques qui prend en charge le gel des E/S.

Le processus de l'instantané continue en arrière-plan tandis que la création de l'instantané est rapide à exécuter et capture un instant donné. Par conséquent, les volumes à sauvegarder ne seront démontés qu'une poignée de secondes. Comme le créneau de sauvegarde est aussi petit que possible, la durée de l'interruption est prévisible et peut être planifiée.

---

[4] https://access.redhat.com/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Storage_Administration_Guide/xfsfreeze.html

Réalisation de sauvegardes multi-volume

Dans certains cas, vous pouvez sauvegarder des données sur plusieurs volumes Amazon EBS grâce au gestionnaire de volumes logiques afin d'augmenter le débit potentiel. Quand vous utilisez un gestionnaire de volumes logiques (mdadm ou LVM, par exemple), il est important d'exécuter la sauvegarde depuis la couche du gestionnaire de volumes plutôt que des volumes EBS sous-jacents. Cela permet de garantir que toutes les métadonnées et les volumes sous-composants sont cohérents.

Il existe différentes façons d'effectuer cela. Par exemple, vous pouvez utiliser le script créé par alestic.com[5]. Les tampons de mémoire doivent être vidés sur le disque ; l'E/S du système de fichiers sur le disque doit être arrêté et un instantané doit être exécuté simultanément pour tous les volumes composant l'ensemble RAID. Après l'exécution de l'instantané pour les volumes (généralement une seconde ou deux), le système de fichiers peut poursuivre ses opérations. Les instantanés doivent être balisés pour que vous puissiez les gérer ensemble pendant une restauration.

Vous pouvez également exécuter ces sauvegardes depuis le gestionnaire de volumes logiques ou au niveau du système de fichiers. Dans ces cas-là, l'utilisation d'un agent de sauvegarde traditionnel permet de sauvegarder les données sur le réseau. De nombreuses solutions de sauvegarde basées sur un agent sont disponibles sur Internet et sur AWS Marketplace.[6] Rappelez-vous qu'un logiciel de sauvegarde basé sur un agent nécessite un nom de serveur cohérent et une adresse IP. Par conséquent, l'utilisation de ces outils combinés aux instances déployées dans un virtual private cloud (VPC)[7] Amazon est la meilleure méthode pour garantir la fiabilité.

Une autre approche consiste à créer une réplique des volumes du système principal qui existent sur un seul gros volume. Cela simplifie le processus de sauvegarde, car un seul gros volume doit être sauvegardé, et la sauvegarde ne s'effectue pas sur le système principal. Toutefois, vous devez déjà déterminer si le volume simple est suffisant pour la sauvegarde et que la taille maximale du volume est appropriée à l'application.

---

[5] https://github.com/alestic/ec2-consistent-snapshot

[6] https://aws.amazon.com/marketplace/

[7] http://aws.amazon.com/vpc/

Approches de la sauvegarde des bases de données

AWS propose de nombreuses options pour les bases de données. Vous pouvez exécuter votre propre base de données sur une instance EC2 ou utiliser une des options de bases de données gérées proposées par Amazon Relational Database Service [8](Amazon RDS). Si vous exécutez votre propre base de données sur une instance EC2, vous pouvez sauvegarder les données sur des fichiers à l'aide des outils natifs (par exemple, MySQL[9], Oracle[10], MSSQL[11], PostgreSQL[12]) ou créer un instantané des volumes contenant les données en utilisant l'une des méthodes décrites dans « Protection basée sur l'instantané EBS ».

Utilisation des sauvegardes des réplicas de bases de données

Pour les bases de données créées sur des ensembles RAID de volumes Amazon EBS, vous pouvez supprimer la charge des sauvegardes sur la base de données principale en créant un réplica en lecture de la base de données. Il s'agit d'une copie à jour de la base de données qui s'exécute sur une instance Amazon EC2 distincte. L'instance de base de données réplica peut être créée en utilisant plusieurs disques similaires à la source, ou les données peuvent être consolidées sur un seul volume EBS. Vous pouvez alors utiliser l'une des procédures décrites dans « Protection basée sur l'instantané EBS » pour faire des instantanés des volumes EBS. Cette approche est souvent utilisée pour des bases de données importantes qui doivent fonctionner 24 h/24, 7 j/7. Dans ce cas, le créneau de sauvegarde est trop long et la base de données de production ne peut pas être désactivée pendant des périodes aussi longues.

Utilisation d'Amazon RDS pour les sauvegardes

Amazon RDS comporte des fonctions pour automatiser les sauvegardes des bases de données. Amazon RDS crée un instantané du volume de stockage de votre instance de base de données. L'intégralité de l'instance DB est sauvegardée, non pas seulement les bases de données.

---

[8] https://aws.amazon.com/rds/

[9] http://dev.mysql.com/doc/refman/5.7/en/backup-and-recovery.html

[10] http://docs.oracle.com/cd/E11882_01/backup.112/e10642/rcmbckba.htm#BRADV8003

[11] http://msdn.microsoft.com/en-us/library/ms187510.aspx

[12] http://www.postgresql.org/docs/9.3/static/backup.html

Amazon RDS fournit deux méthodes différentes de sauvegarde et de restauration des instances DB :

- **Les sauvegardes automatisées** garantissent une restauration à un instant dans le passé de votre instance DB. Les sauvegardes automatiques sont activées par défaut lorsque vous créez une nouvelle instance DB. Amazon RDS exécute une sauvegarde complète quotidienne de vos données pendant un créneau que vous définissez lorsque vous créez l'instance DB. Vous pouvez configurer une période de rétention de 35 jours maximum pour les sauvegardes automatiques. Amazon RDS utilise ces sauvegardes régulières de données périodiques avec vos journaux de transactions pour vous permettre de restaurer votre instance DB à tout instant de votre période de rétention, jusqu'à la valeur `LatestRestorableTime` (généralement les cinq dernières minutes). Pour trouver la date de restauration la plus récente pour vos instances de base de données, vous pouvez utiliser l'appel d'API `DescribeDBInstances` ou chercher dans l'onglet **Description** la base de données dans AWS Management Console[AA1].

  Quand vous lancez une restauration à un instant dans le passé, les journaux de transactions sont appliqués à la sauvegarde quotidienne la plus appropriée afin de restaurer votre instance DB au moment que vous aurez demandé.

- **Les snapshots DB** sont des sauvegardes initiées par l'utilisateur qui permettent de sauvegarder votre instance DB dans un état connu aussi fréquemment que vous le souhaitez, puis de la restaurer dans cet état spécifique à tout moment. Vous pouvez utiliser AWS Management Console ou l'appel d'API `CreateDBSnapshot` pour créer des snapshots DB. Ces instantanés possèdent une rétention illimitée. Ils sont conservés jusqu'à ce que vous utilisiez la console ou l'appel d'API `DeleteDBSnapshot` pour les supprimer de manière explicite.

Lorsque vous restaurez une base de données à un instant donné ou à partir d'un snapshot DB, une nouvelle instance de base de données sera créée avec un nouveau point de terminaison. De cette manière, vous pouvez créer plusieurs instances de base de données à partir d'un snapshot DB spécifique ou d'un instant donné.

Vous pouvez utiliser AWS Management Console ou un appel `DeleteDBInstance` pour supprimer l'ancienne instance de base de données.

Utilisation d'une AMI pour sauvegarder des instances EC2

AWS stocke des images système dans ce qui s'appelle Amazon Machine Images (AMI). Ces images sont le template du volume racine requis pour lancer une instance. Vous pouvez utiliser AWS Management Console ou la commande de l'interface de ligne de commande `aws ec2 create-image` pour sauvegarder le volume racine comme une AMI.

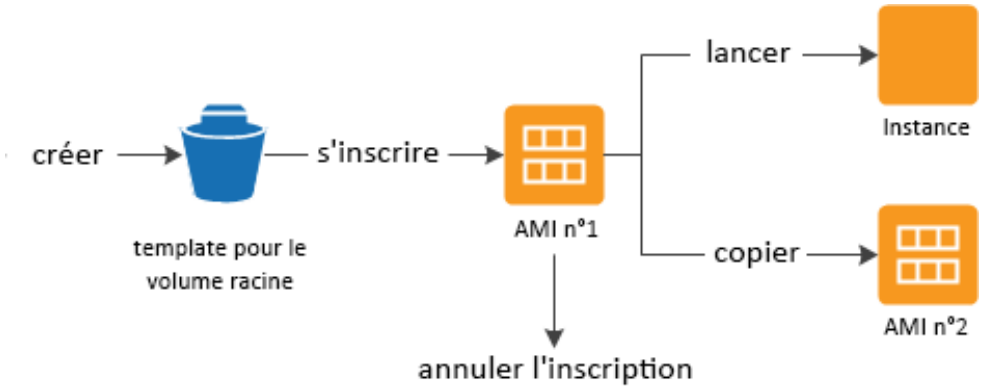
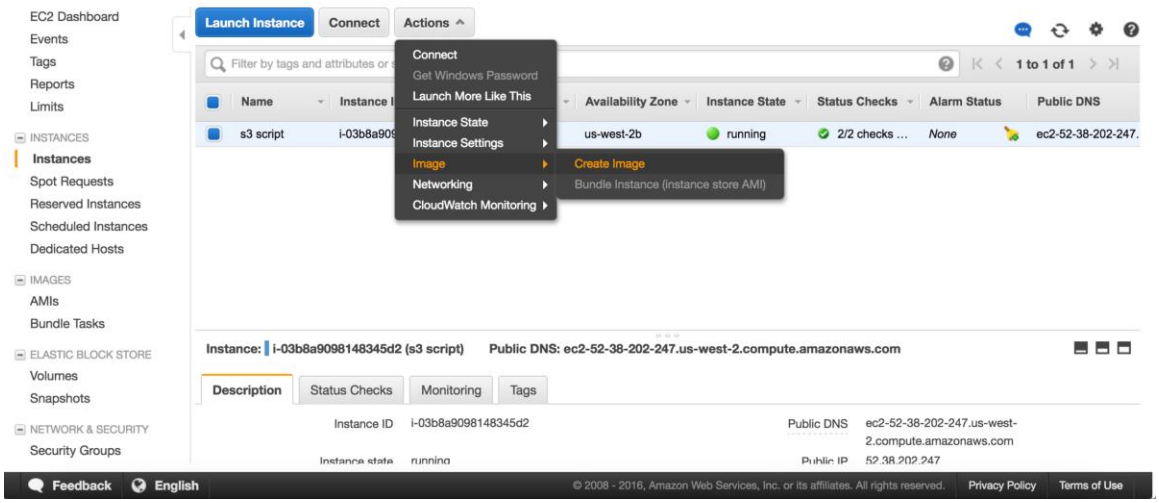


Figure3 : Utilisation d'une AMI pour sauvegarder et lancer une instance

Quand vous enregistrez une AMI, elle est stockée dans votre compte à l'aide des instantanés Amazon EBS. Ces instantanés résident dans Amazon S3 de façon à offrir une durabilité élevée.



[PAL2]

Figure4 : Utilisation de la console EC2 pour créer une image de machine

Une fois que vous avez créé une AMI de votre instance Amazon EC2, vous pouvez utiliser l'AMI pour recréer l'instance ou lancer davantage de copies de l'instance.

Vous pouvez également copier les AMI à partir d'une région dans une autre région pour la migration d'applications ou la reprise après sinistre.

Infrastructure sur site vers AWS

Ce scénario décrit un environnement de charge de travail sans aucun composant dans le cloud. Toutes les ressources, incluant les serveurs Web, les serveurs d'application, les serveurs de supervision, les bases de données, les services Active Directory, etc., sont hébergées sur le centre de données du client ou en colocation.

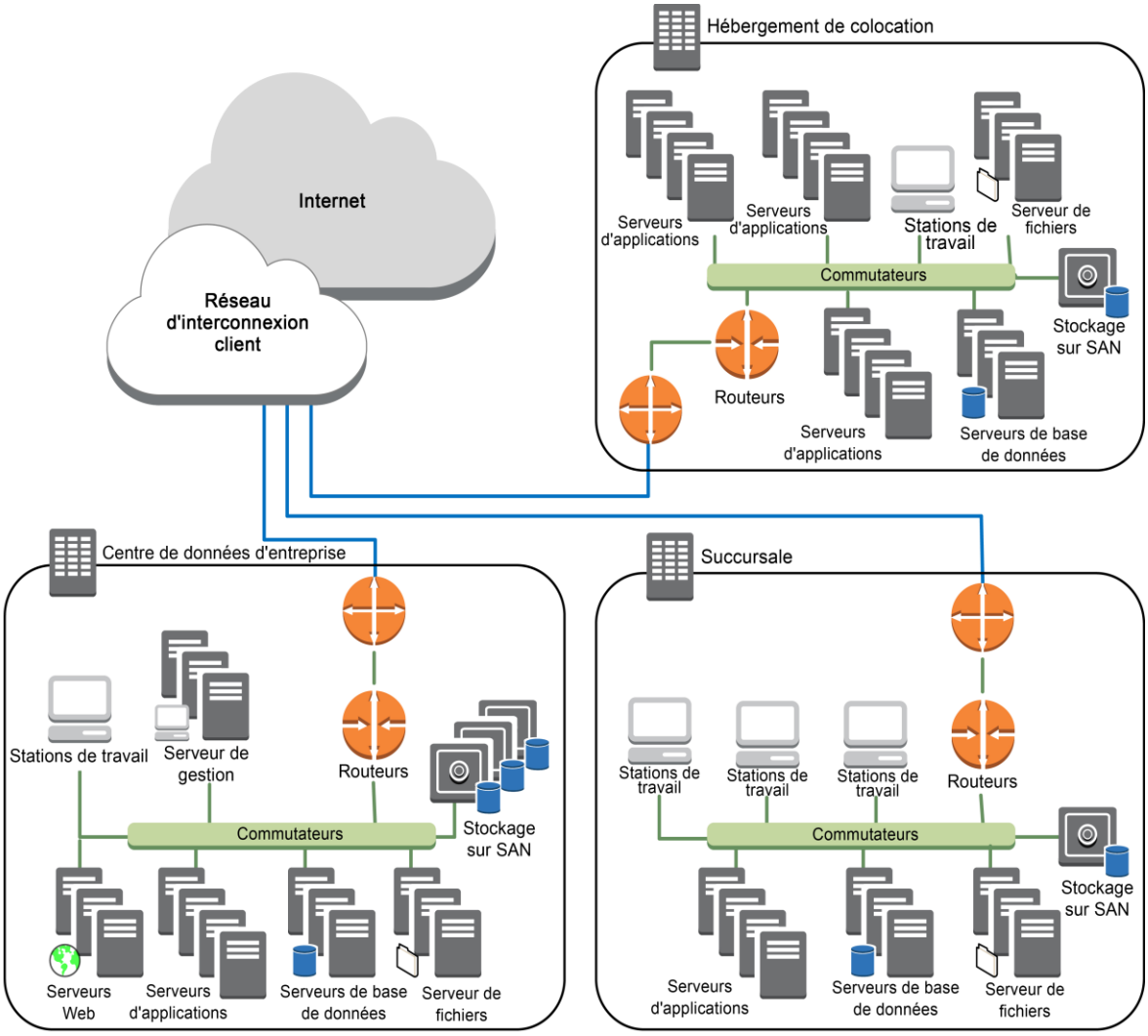


Figure5 : Environnement sur site

En utilisant les services de stockage AWS de ce scénario, vous pouvez vous concentrer sur les tâches de sauvegarde et d'archivage. Vous n'avez pas à vous soucier du dimensionnement du stockage ou de la capacité de l'infrastructure pour accomplir la tâche de sauvegarde.

Amazon S3 et Amazon Glacier sont des API basées en mode natif et disponibles via Internet. Elles permettent aux fournisseurs de logiciel de sauvegarde d'intégrer directement leurs applications aux solutions de stockage AWS, comme indiqué dans la figure suivante.

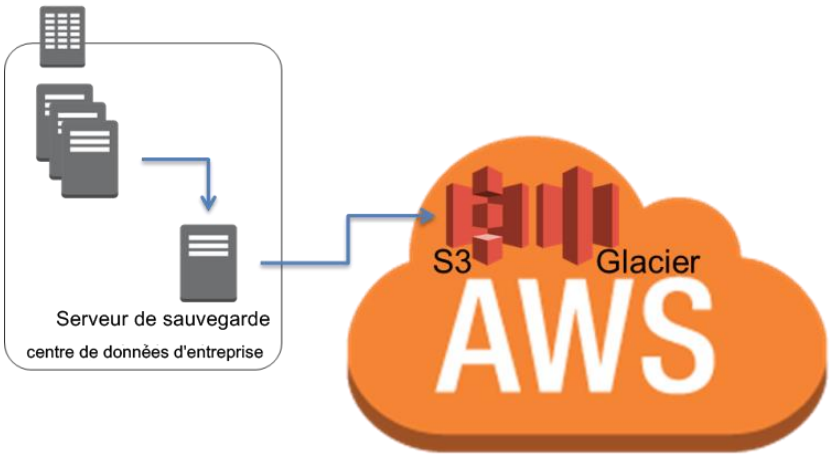


Figure6 : Connecteur de sauvegarde vers Amazon S3 ou Amazon Glacier

Dans ce scénario, le logiciel de sauvegarde et d'archivage communique directement avec AWS à travers les API. Comme le logiciel de sauvegarde est compatible avec AWS, il peut sauvegarder les données depuis les serveurs sur site directement dans Amazon S3 ou Amazon Glacier.

Si votre logiciel de sauvegarde existant ne prend pas en charge le cloud AWS en mode natif, vous pouvez utiliser nos produits AWS Storage Gateway. Le service AWS Storage Gateway[13] est une application virtuelle qui fournit une intégration continue et sécurisée entre votre centre de données et l'infrastructure de stockage AWS. Ce service vous permet de stocker vos données, en toute sécurité, dans le cloud AWS, pour un stockage évolutif et économique. AWS Storage Gateway prend en charge les protocoles de stockage standard du secteur, qui fonctionnent avec vos applications existantes tout en assurant le stockage sécurisé de l'ensemble de vos données chiffrées dans Amazon S3 ou Amazon Glacier.

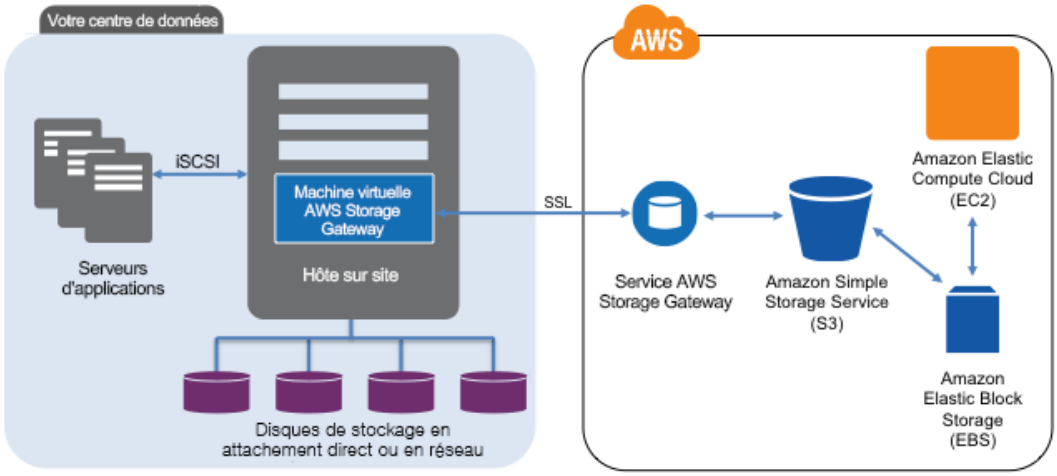


Figure7 : Connexion des ressources sur site au stockage AWS

AWS Storage Gateway propose les configurations suivantes :

- **Passerelles de volume :** Les passerelles de volume offrent des volumes de stockage dans le cloud que vous pouvez monter comme des périphériques iSCSI (Internet Small Computer System Interface) à partir de vos serveurs d'applications sur site. La passerelle propose les configurations de volume suivantes :

[13] http://aws.amazon.com/storagegateway/

  - **Volumes mis en cache sur la passerelle :** Vous pouvez stocker vos données primaires dans Amazon S3 et conserver en local vos données fréquemment consultées. Les volumes mis en cache sur Gateway vous permettent de réaliser d'importantes économies sur le stockage de vos données primaires, puisque vous n'avez plus forcément besoin d'ajuster votre capacité de stockage sur site. De plus, vous conservez un accès à faible latence à vos données fréquemment consultées.
  - **Volumes stockés sur la passerelle :** Si vous avez besoin d'un accès avec un faible temps de latence à l'ensemble de vos données, vous pouvez configurer votre passerelle de données sur site de façon à stocker vos données primaires en local, puis sauvegarder de manière asynchrone des instantanés ponctuels de ces données dans Amazon S3. Les volumes stockés sur Gateway fournissent des sauvegardes hors site durables et peu coûteuses, que vous pouvez récupérer localement ou à partir d'Amazon EC2.
- **Gateway-virtual tape library (gateway-VTL) :** Avec Gateway-VTL, vous pouvez disposer d'une collection illimitée de bandes virtuelles. Chaque bande virtuelle peut être stockée dans une bibliothèque de bandes virtuelles (Virtual Tape Library) sauvegardée par Amazon S3 ou une étagère de bandes virtuelles (Virtual Tape Shelf) sauvegardée par Amazon Glacier. La bibliothèque de bandes virtuelles expose une interface iSCSI conforme aux normes du secteur, qui fournit à votre application de sauvegarde un accès en ligne aux bandes virtuelles. Dès lors que vous n'avez plus besoin d'un accès fréquent ou immédiat aux données contenues dans une bande virtuelle, vous pouvez utiliser votre application de sauvegarde pour déplacer cette bande de sa bibliothèque de bandes virtuelles vers votre étagère de bandes virtuelles, afin de réduire encore vos coûts de stockage.

Ces passerelles agissent comme des périphériques Plug-and-Play fournissant des périphériques iSCSI standard pouvant être intégrés à votre infrastructure de sauvegarde ou d'archivage. Vous pouvez utiliser les périphériques de disque iSCSI comme pools de stockage de votre logiciel de sauvegarde ou la passerelle VTL pour transférer la sauvegarde sur bande ou archiver directement dans Amazon S3 ou Amazon Glacier.

Avec cette solution, vos sauvegardes et archives sont automatiquement effectuées hors site (pour des raisons de conformité) et stockées sur un support durable, éliminant ainsi la complexité et les risques de sécurité de la gestion de bandes hors site.

Environnements hybrides

Les deux déploiements d'infrastructure traités jusque-là, résidant dans le cloud et sur site, peuvent être associés dans un scénario hybride dont l'environnement de charge de travail possède des composants d'infrastructure sur site et sur AWS. Les ressources, incluant les serveurs Web, les serveurs d'application, les serveurs de supervision, les bases de données, les services Active Directory, etc., sont hébergées sur le centre de données du client ou sur AWS. Les applications exécutées dans le cloud AWS sont connectées à celles qui sont exécutées sur site.

Ce scénario est en passe de devenir courant pour les charges de travail de l'entreprise. De nombreuses entreprises possèdent leurs propres centres de données et utilisent AWS pour augmenter leur capacité. Leurs centres de données sont souvent connectés au réseau AWS par des liens vers un réseau haute capacité. Par exemple, avec AWS Direct Connect[14], vous pouvez établir une connectivité dédiée privée de vos locaux vers AWS. [PAL3]Cela fournit la bande passante et une latence fiable pour télécharger des données vers le Cloud dans le cadre de la protection des données ainsi qu'une performance et une latence fiables pour les charges de travail hybrides.

[14] http://aws.amazon.com/directconnect/

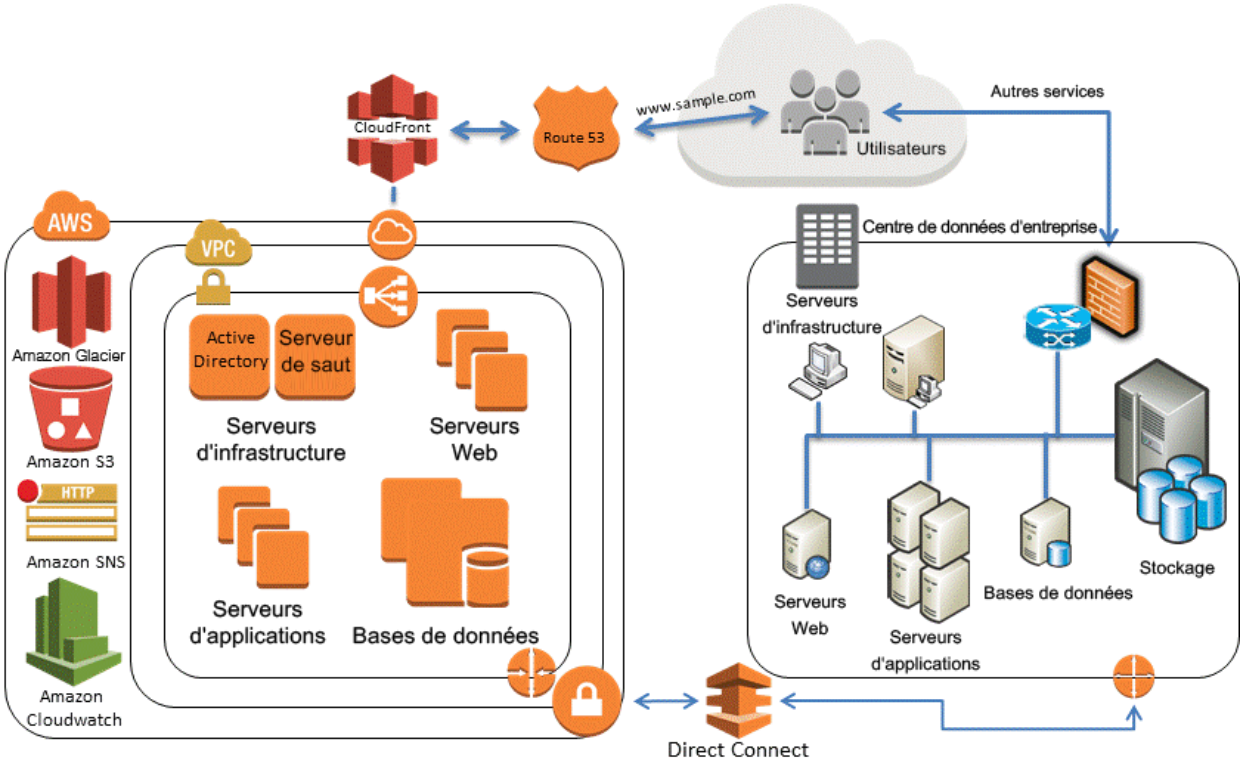


Figure8 : Un scénario d'une infrastructure hybride

Les solutions de protection des données bien conçues utilisent une combinaison des méthodes décrites dans les solutions fondées sur le cloud et sur site.

Sauvegarde d'applications basées sur AWS vers votre centre de données

Si vous possédez déjà une infrastructure qui sauvegarde les données sur vos serveurs sur site, son extension vers vos ressources AWS est très simple par le biais d'une connexion VPN ou d'AWS Direct Connect. Vous pouvez installer l'agent de sauvegarde sur les instances Amazon EC2 et les sauvegarder selon vos stratégies de protection des données.

Migration de la gestion des sauvegardes vers le Cloud pour plus de disponibilité

En fonction de l'architecture de votre sauvegarde, vous pouvez disposer d'un serveur de sauvegarde principal et d'un ou de plusieurs serveurs médias ou de stockage situés sur site avec les services protégés. Dans ce cas, vous pouvez décider de déplacer le serveur de sauvegarde principal sur une instance Amazon EC2 pour le protéger des sinistres sur site et disposer ainsi d'une infrastructure de sauvegarde hautement disponible.

Pour gérer les flux de données sauvegardées, vous pouvez également créer un ou plusieurs serveurs médias sur les instances Amazon EC2. Les serveurs médias proches des instances Amazon EC2 vous feront économiser de l'argent au niveau du transfert de données Internet et ils amélioreront la performance générale en matière de sauvegarde et de récupération lors de la sauvegarde vers S3 ou Amazon Glacier.

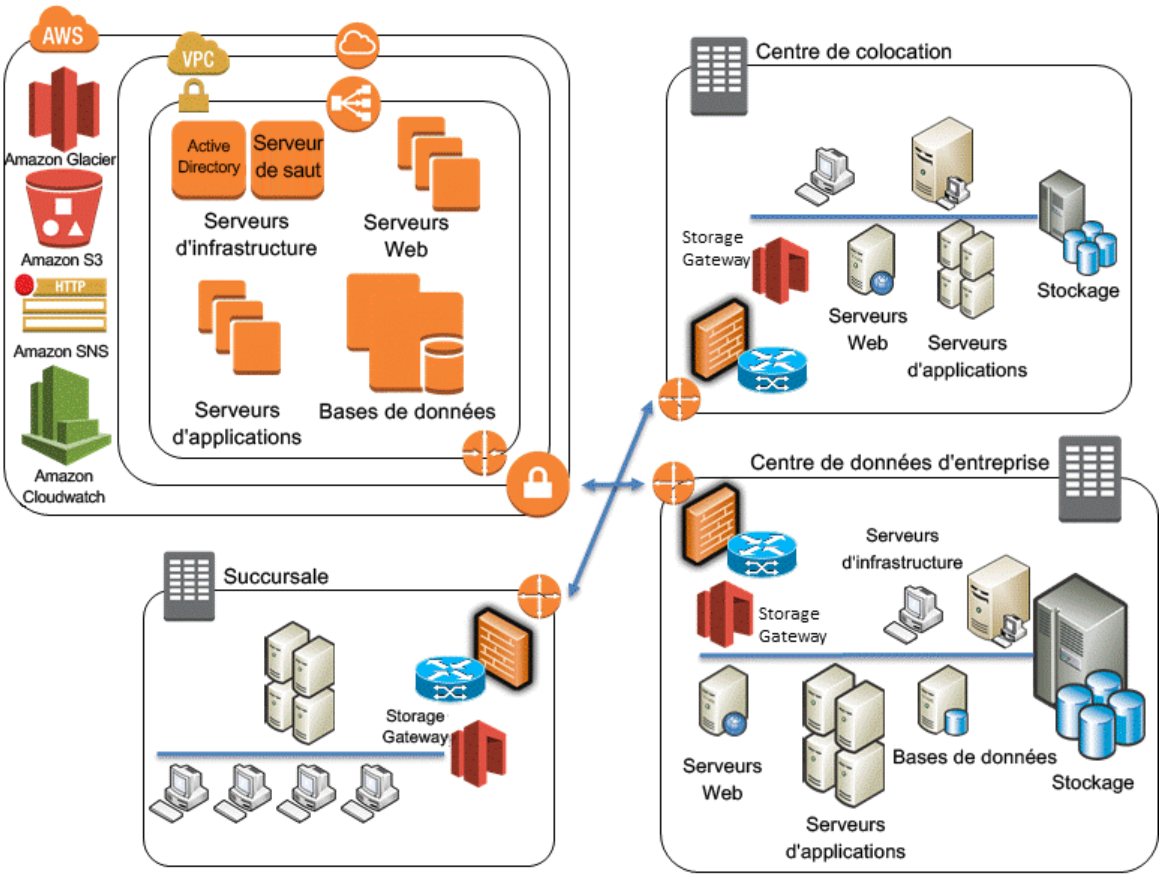


Figure9 : Utilisation des passerelles dans un scénario hybride

Exemple de scénario hybride

Supposez que vous êtes responsable de la gestion d'un environnement où vous devez sauvegarder des instances Amazon EC2, des serveurs autonomes, des machines virtuelles et des bases de données. Cet environnement comporte 1 000 serveurs, et les sauvegardes concernent le système d'exploitation, les données de fichiers, les images de machine virtuelle et les bases de données. Il existe 20 bases de données (composées de MySQL, Microsoft SQL Server et Oracle) à sauvegarder.

Votre logiciel de sauvegarde possède des agents qui sauvegardent les systèmes d'exploitation, les images de machine virtuelle, les volumes de données, les bases de données SQL Server et les bases de données Oracle (avec RMAN). Pour les applications comme MySQL pour lesquelles votre logiciel de stockage ne possède pas d'agent, vous pouvez utiliser l'utilitaire client mysqldump afin de créer un fichier de vidage de la base de données sur le disque où les agents de sauvegarde standards peuvent alors protéger les données.

Pour protéger cet environnement ci-dessus, votre logiciel de sauvegarde tiers dispose probablement d'un serveur catalogue global ou serveur principal qui contrôle les activités de sauvegarde, d'archivage et de restauration, ainsi que plusieurs serveurs médias qui sont connectés au stockage sur disque, aux lecteurs de bande LTO et aux services de stockage AWS.

La méthode la plus simple d'améliorer votre solution de sauvegarde avec des services de stockage AWS est de profiter du support de votre fournisseur de sauvegarde pour Amazon S3 ou Amazon Glacier. Nous vous suggérons de travailler avec votre fournisseur pour comprendre leurs options d'intégration et de connecteur. Pour une liste de fournisseurs de logiciel de sauvegarde qui travaillent avec AWS, consultez notre répertoire de partenaires[15].

Si votre logiciel de sauvegarde existant ne prend pas en charge le stockage dans le cloud en mode natif des sauvegardes ou archives, vous pouvez utiliser un périphérique de passerelle de stockage, comme un pont, entre le logiciel de sauvegarde et Amazon S3 ou Amazon Glacier.

---

[15] http://www.aws-partner-directory.com/PartnerDirectory/PartnerSearch?type=ISV

Il existe plusieurs solutions de passerelles tierces. Vous pouvez aussi utiliser les périphériques virtuels AWS Storage Gateway pour combler ce fossé, car ils utilisent des techniques génériques comme des volumes iSCSI et des bibliothèques de bandes virtuelles (VTL). Cette configuration nécessite un hyperviseur pris en charge (VMware ou Microsoft Hyper-V) et un stockage local pour héberger le périphérique.

Archivage de données avec AWS

Lorsque vous devez préserver des données pour des raisons de conformité ou commerciales, vous les archivez. Contrairement aux sauvegardes, qui sont généralement effectuées pour garder une copie des données de production pour une courte durée afin de les récupérer après un endommagement ou une perte des données, l'archivage conserve toutes les copies des données jusqu'à l'expiration de la stratégie de rétention.

De bonnes archives répondent aux critères suivants :

- Durabilité des données pour une intégrité à long terme
- Sécurité des données
- Facilité de récupération
- Faible coût

Des magasins de données immuables peuvent constituer une autre condition de réglementation ou de conformité.

Amazon Glacier propose des archives à un faible coût, un chiffrement natif des données au repos, une durabilité de 99,999999999 % (lisez « onze-neuf ») et une capacité illimitée.

Amazon S3 Standard-Infrequent Access convient plus particulièrement pour les cas d'utilisation qui nécessitent la récupération rapide des données. Amazon Glacier constitue un choix idéal pour les cas d'utilisation dans lesquels les données sont utilisées de manière occasionnelle et les délais de récupération de plusieurs heures restent acceptables.

Les objets peuvent être hiérarchisés dans Amazon Glacier avec les règles de cycle de vie de S3 ou l'API d'Amazon Glacier. La fonction de verrouillage de coffre Amazon Glacier vous permet de déployer et d'appliquer facilement les contrôles de conformité pour les coffres Amazon Glacier individuels avec une stratégie de verrouillage de coffre. Vous pouvez spécifier des contrôles tels que « WROM » dans une stratégie de verrouillage de coffre, puis verrouiller la stratégie pour éviter de futures modifications. Pour en savoir plus, consultez Amazon Glacier.

Sécurisation des données de sauvegarde avec AWS

La sécurité des données est une préoccupation courante. AWS ne badine pas avec la sécurité. Il s'agit du fondement de chaque service que nous lançons. Les services de stockage comme Amazon S3 offrent des mesures de contrôle d'accès et de chiffrement avancées, que les données soient au repos ou en transit.. Tous les points de terminaison API Amazon S3 et Amazon Glacier prennent en charge le chiffrement SSL pour les données en transit. Amazon Glacier chiffre toutes les données au repos par défaut. Avec Amazon S3, les clients peuvent choisir le chiffrement côté serveur pour les objets au repos en laissant AWS gérer les clés de chiffrement, en fournissant leurs propres clés lorsqu'ils chargent un objet ou en utilisant l'intégration d'AWS Key Management Service (AWS KMS) [16] pour les clés de chiffrement. Sinon, les clients peuvent toujours chiffrer leurs données avant de les charger vers AWS. Pour plus d'informations, consultez Amazon Web Services: Overview of Security Processes.

[16] http://docs.aws.amazon.com/AmazonS3/latest/dev/UsingKMSEncryption.html

Conclusion

Gartner classe AWS au rang de leader en termes de services de stockage dans le cloud public[17]. AWS est bien placé pour aider les entreprises à basculer leurs charges de travail sur les plateformes de cloud qui offrent une nouvelle génération de stockage. AWS fournit des solutions économiques et évolutives pour aider les entreprises à équilibrer leurs besoins de sauvegarde et d'archivage. Ces services s'intègrent parfaitement aux technologies que vous utilisez aujourd'hui.

Collaborateurs

Les personnes qui suivent ont participé à l'élaboration de ce document :

- Pawan Agnihotri, Architecte de solutions, Amazon Web Services
- Lee Kear, architecte de solutions, Amazon Web Services
- Peter Levett, Architecte de solutions, Amazon Web Services

Révisions de documents

Mis à jour en mai 2016

[17] http://www.gartner.com/technology/reprints.do?id=1-1WWKTQ3&ct=140709&st=sb