



# Commencez

## GenAI

NetApp  
October 06, 2025

# Sommaire

Commencez . . . . .	1
Démarrage rapide pour les connecteurs GenAI . . . . .	1
Exigences relatives au connecteur GenAI . . . . .	2
Exigences GenAI de base . . . . .	2
Configuration requise pour NetApp Connector pour Amazon Q Business . . . . .	2
Identifier les sources de données à ajouter à un connecteur . . . . .	3
Nombre maximum de sources de données . . . . .	3
Emplacement des sources de données . . . . .	3
Protocoles pris en charge . . . . .	3
Formats de fichiers de source de données pris en charge . . . . .	4
Déployez l'infrastructure GenAI . . . . .	4
Détails de l'infrastructure . . . . .	5
Déployez l'infrastructure GenAI . . . . .	6

# Commencez

## Démarrage rapide pour les connecteurs GenAI

Commencez à créer un connecteur NetApp pour Amazon Q Business à l'aide des données de votre organisation qui existent sur les systèmes de fichiers Amazon FSx pour NetApp ONTAP. Après avoir créé un connecteur, les utilisateurs finaux peuvent accéder à l'assistant Amazon Q Business pour obtenir des réponses axées sur l'organisation à leurs questions.

1

### Connectez-vous à Workload Factory

Vous aurez besoin de "[créer un compte avec Workload Factory](#)" et connectez-vous en utilisant l'un des "[expériences de la console](#)".

2

### Configurez votre environnement pour répondre aux exigences GenAI

Vous aurez besoin d'informations d'identification AWS pour déployer l'infrastructure AWS, d'un système de fichiers FSX for ONTAP déployé et découvert, de la liste des sources de données que vous souhaitez intégrer à votre connecteur, de l'accès à l'application Amazon Q Business et bien plus encore.

["En savoir plus sur les exigences GenAI".](#)

3

### Identifiez le système de fichiers FSX pour ONTAP qui contient les sources de données

Les sources de données que vous intégrerez dans votre connecteur peuvent être situées dans un seul système de fichiers FSX pour ONTAP ou dans plusieurs systèmes de fichiers FSX pour ONTAP. Si ces systèmes se trouvent dans des VPC différents, ils doivent être accessibles au sein du même réseau, ou les VPC doivent être associés et utiliser le même compte région et AWS que le moteur d'IA.

["Découvrez comment identifier les sources de données".](#)

4

### Déployez l'infrastructure GenAI

Lancez l'assistant de déploiement d'infrastructure pour déployer l'infrastructure GenAI dans votre environnement AWS. Ce processus déploie une instance EC2 pour le moteur NetApp GenAI et un volume sur un système de fichiers FSX pour ONTAP afin de contenir les bases de données du moteur NetApp ai. Le volume est utilisé pour stocker des informations sur le connecteur.

["Découvrez comment déployer l'infrastructure GenAI".](#)

### Et la suite

Vous pouvez maintenant créer un connecteur pour Amazon Q Business afin de fournir des réponses axées sur l'organisation aux utilisateurs finaux.

# Exigences relatives au connecteur GenAI

Assurez-vous que Workload Factory et AWS sont correctement configurés avant de créer un connecteur NetApp pour Amazon Q Business.

## Exigences GenAI de base

GenAI présente des exigences générales que votre environnement doit satisfaire avant de commencer.

### Connexion et compte Workload Factory

Vous aurez besoin de "[créer un compte avec Workload Factory](#)" et connectez-vous en utilisant l'un des "[expériences de la console](#)".

### Identifiants et autorisations AWS

Vous devez ajouter les informations d'identification AWS à Workload Factory avec des autorisations de lecture/écriture, ce qui signifie que vous utiliserez Workload Factory en mode *lecture/écriture* pour GenAI.

Les autorisations en mode *de base* et en mode *lecture seule* ne sont pas prises en charge pour le moment.

Lors de la configuration de vos identifiants, le fait de sélectionner les autorisations comme indiqué ci-dessous vous permet de bénéficier d'un accès complet pour gérer les systèmes de fichiers FSX pour ONTAP, ainsi que pour déployer et gérer l'instance EC2 GenAI et d'autres ressources AWS nécessaires à votre base de connaissances et à votre chatbot.

["Découvrez comment ajouter des informations d'identification AWS à Workload Factory"](#)

## Configuration requise pour NetApp Connector pour Amazon Q Business

Assurez-vous que votre environnement répond aux exigences spécifiques suivantes pour les connecteurs Amazon Q Business.

### Application Amazon Q Business

Vous devez créer une application Amazon Q Business ou utiliser une application existante.

- Assurez-vous que l'application existe dans l'une de vos régions AWS.
- Assurez-vous que vous avez "[créé un index](#)" pour l'application.
- Assurez-vous que l'application n'est pas en état d'échec.

### FSX pour le système de fichiers ONTAP

Vous avez besoin d'au moins un système de fichiers FSX pour ONTAP :

- Un système de fichiers sera utilisé (ou créé, s'il n'existe pas) par le moteur NetApp GenAI pour stocker des informations sur le connecteur.

Ce système de fichiers FSX pour ONTAP doit utiliser des volumes FlexVol. Les volumes FlexGroup ne sont pas pris en charge.

- Un ou plusieurs systèmes de fichiers contiennent les sources de données que vous allez ajouter à votre connecteur.

Un système de fichiers FSX pour ONTAP peut être utilisé pour ces deux raisons ou vous pouvez utiliser plusieurs systèmes de fichiers FSX pour ONTAP.

- Vous devez connaître la région AWS, le VPC et le sous-réseau où réside le système de fichiers AWS FSX pour ONTAP.
- Vous devrez tenir compte des paires clé/valeur que vous souhaitez appliquer aux ressources AWS faisant partie de ce déploiement ( facultatif).
- Vous devrez connaître les informations relatives aux paires de clés qui vous permettront de vous connecter en toute sécurité à l'instance du moteur NetApp ai.

"Découvrez comment déployer et gérer les systèmes de fichiers FSX pour ONTAP"

## Identifier les sources de données à ajouter à un connecteur

Identifiez ou créez les documents (sources de données) qui résident dans votre système de fichiers FSX pour ONTAP que vous allez intégrer à votre connecteur. Ces sources de données permettent à Amazon Q Business de fournir des réponses précises et personnalisées aux questions des utilisateurs en fonction des données pertinentes pour votre entreprise.

### Nombre maximum de sources de données

Le nombre maximum de sources de données prises en charge est de 10.

### Emplacement des sources de données

Les sources de données peuvent être stockées dans un seul volume, ou dans un dossier au sein d'un volume, sur un partage SMB ou une exportation NFS sur un système de fichiers Amazon FSX pour NetApp ONTAP. Les sources de données peuvent également être stockées dans des volumes Amazon FSX pour NetApp ONTAP dans une relation de protection des données NetApp SnapMirror.

Vous ne pouvez pas sélectionner des documents individuels au sein d'un volume ou d'un dossier. Par conséquent, vous devez vous assurer que chaque volume ou dossier contenant des sources de données ne contient pas de documents externes qui ne doivent pas être intégrés à votre base de connaissances.

Vous pouvez ajouter plusieurs sources de données dans chaque connecteur, mais elles doivent résider sur des systèmes de fichiers FSX pour ONTAP accessibles depuis votre compte AWS.

La taille de fichier maximale pour chaque source de données est de 50 Mo.

### Protocoles pris en charge

Les connecteurs prennent en charge les données des volumes qui utilisent les protocoles NFS ou SMB/CIFS. Lorsque vous sélectionnez des fichiers stockés à l'aide du protocole SMB, vous devez entrer les informations Active Directory pour que le connecteur puisse accéder aux fichiers de ces volumes. Ceci inclut le domaine Active Directory, l'adresse IP, le nom d'utilisateur et le mot de passe.

Lorsque vous stockez votre source de données sur un partage (fichier ou répertoire) accessible via SMB, les données sont uniquement accessibles aux utilisateurs ou groupes de chatbot qui disposent des autorisations d'accès à ce partage. Lorsque cette « fonctionnalité de reconnaissance des autorisations » est activée, le système d'IA compare le courrier électronique de l'utilisateur dans auth0 aux utilisateurs autorisés à afficher ou utiliser les fichiers sur le partage SMB. Le chatbot fournira des réponses en fonction des autorisations utilisateur pour les fichiers intégrés.

Par exemple, si vous avez intégré 10 fichiers (sources de données) dans votre connecteur et que 2 fichiers sont des fichiers de ressources humaines qui contiennent des informations restreintes, seuls les utilisateurs de chatbot qui sont authentifiés pour accéder à ces 2 fichiers recevront des réponses du chatbot qui incluent des données de ces fichiers.



Lorsque vous ajoutez des sources de données à Amazon Q Business Connector, seules les autorisations utilisateur s'appliquent aux fichiers de source de données. Les autorisations de groupe ne sont pas appliquées.



Si un fichier de votre source de données manque de texte (par exemple, une image sans texte), Amazon Q Business ne l'indexe pas mais consigne une entrée dans les journaux Amazon CloudWatch en notant l'absence de texte.

## Formats de fichiers de source de données pris en charge

Les formats de fichiers de source de données suivants sont actuellement pris en charge avec NetApp Connector pour Amazon Q Business.

Format de fichier	Extension
Fichier de valeurs séparées par des virgules	.csv
JSON et JSONP	.json
Démarque	.md
Microsoft Word	.docx
Texte brut	.txt
Format de document portable	.pdf
Microsoft PowerPoint	.ppt ou .pptx
Langage de balisage hypertexte	.html
Langage de balisage extensible	.xml
XSLT	.xslt
Microsoft Excel	.xls
Format texte riche	.rtf

## Déployez l'infrastructure GenAI

Vous devez déployer l'infrastructure GenAI pour la structure RAG dans votre environnement avant de pouvoir créer des bases de connaissances, des connecteurs et des applications FSX pour ONTAP pour votre entreprise. Les principaux composants de l'infrastructure sont le service Amazon Bedrock, une instance de machine virtuelle pour le moteur NetApp GenAI et un système de fichiers FSX pour ONTAP.

L'infrastructure déployée peut prendre en charge plusieurs bases de connaissances, chatbots et connecteurs. Vous n'aurez donc généralement besoin d'effectuer cette tâche qu'une seule fois.

## Détails de l'infrastructure

Votre déploiement GenAI doit se trouver dans une région AWS sur laquelle Amazon Bedrock est activé.

["Afficher la liste des régions prises en charge"](#)

L'infrastructure comprend les composants suivants.

### Service Amazon Bedrock

Amazon Bedrock est un service entièrement géré qui vous permet d'utiliser les modèles de base (FMS) des plus grandes entreprises d'IA via une seule API. Il fournit également les fonctionnalités nécessaires pour créer des applications d'IA générative sécurisées.

["En savoir plus sur Amazon Bedrock"](#)

### Amazon Q Business

Amazon Q s'appuie sur Amazon Bedrock pour fournir un assistant d'IA générative entièrement géré que vous pouvez utiliser pour répondre à des questions et générer du contenu à partir d'informations issues de vos sources de données.

["En savoir plus sur Amazon Q Business"](#)

### Machine virtuelle pour le moteur NetApp GenAI

Le moteur NetApp GenAI est déployé au cours de ce processus. Il fournit la puissance de traitement permettant d'acquérir les données de vos sources de données, puis d'écrire ces données dans la base de données vectorielle.

### FSX pour le système de fichiers ONTAP

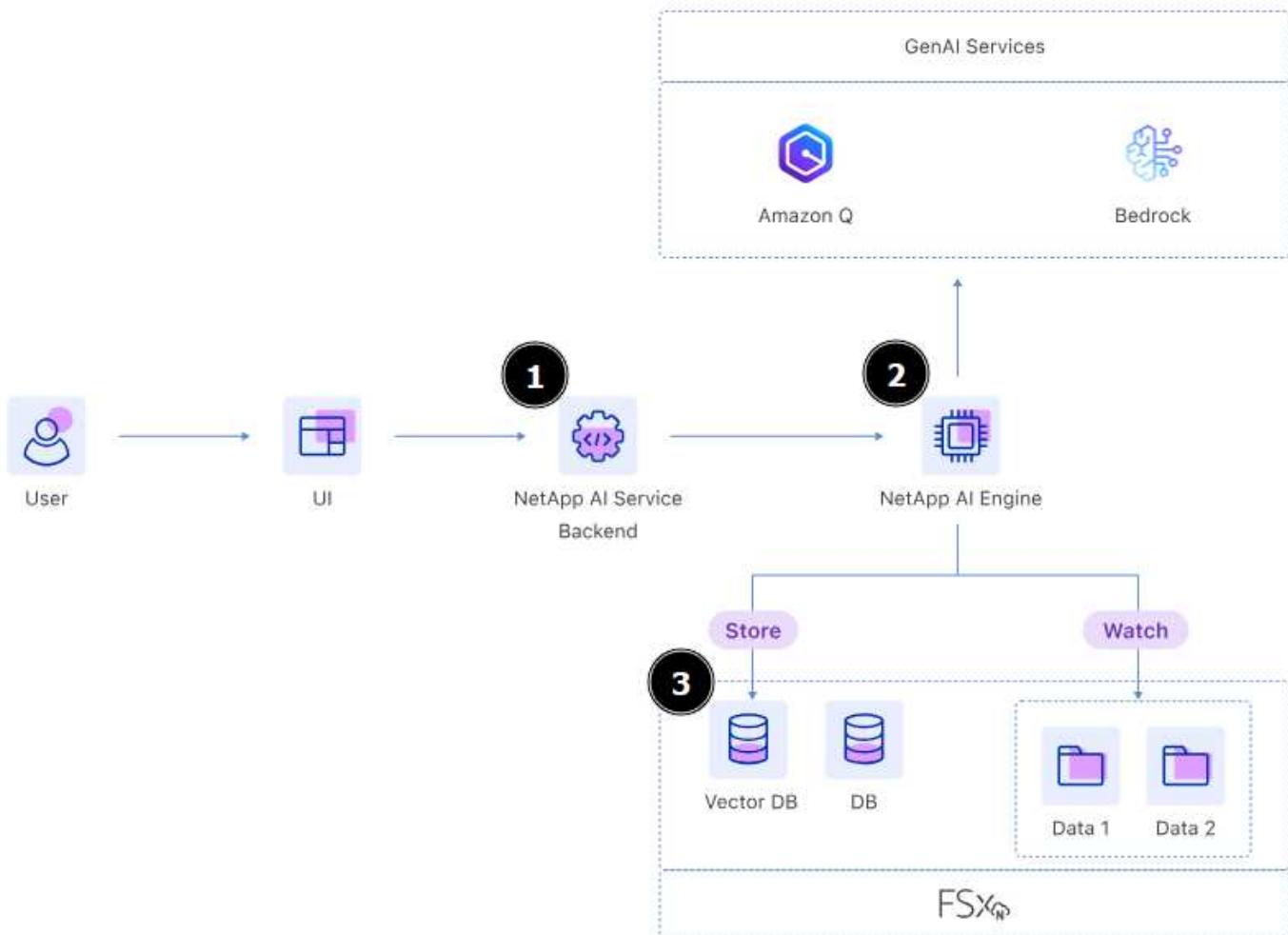
Le système de fichiers FSX pour ONTAP fournit le stockage de votre système GenAI.

Un seul volume est déployé qui contiendra la base de données vectorielle qui stocke les données générées par le modèle de base basé sur vos sources de données.

Les sources de données que vous intégrerez dans votre base de connaissances peuvent résider sur le même système de fichiers FSX pour ONTAP ou sur un autre système.

Le moteur NetApp GenAI surveille ces volumes et interagit avec eux.

L'image suivante montre l'infrastructure GenAI. Les composants numérotés 1, 2 et 3 sont déployés au cours de cette procédure. Les autres éléments doivent être en place avant de commencer le déploiement.



## Déployez l'infrastructure GenAI

Vous devrez saisir vos identifiants AWS et sélectionner le système de fichiers FSX pour ONTAP pour déployer l'infrastructure de génération augmentée de récupération.

### Avant de commencer

Avant de commencer cette procédure, assurez-vous que votre environnement répond aux conditions requises pour les bases de connaissances ou les connecteurs, selon ce que vous choisissez.

- "Exigences de la base de connaissances"
- "Exigences relatives au connecteur"

### Étapes

1. Connectez-vous à Workload Factory en utilisant l'un des "[expériences de la console](#)".
2. Dans la mosaïque workloads d'IA, sélectionnez **déployer et gérer**.
3. Examinez le diagramme d'infrastructure et sélectionnez **Suivant**.
4. Complétez les éléments de la section **paramètres AWS** :
  - a. **Informations d'identification AWS** : sélectionnez ou ajoutez les informations d'identification AWS qui fournissent des autorisations pour déployer les ressources AWS.
  - b. **Emplacement** : sélectionnez une région AWS, un VPC et un sous-réseau.

Le déploiement GenAI doit se trouver dans une région AWS sur laquelle le système Amazon Bedrock est activé. ["Afficher la liste des régions prises en charge"](#)

5. Complétez les éléments de la section **Paramètres de l'infrastructure** :

- Tags** : saisissez les paires clé/valeur de balise que vous souhaitez appliquer à toutes les ressources AWS qui font partie de ce déploiement. Ces balises sont visibles dans la console de gestion AWS et dans la zone d'informations sur l'infrastructure de Workload Factory et peuvent vous aider à suivre les ressources de Workload Factory.

6. Complétez la section **connectivité** :

- Paire de clés** : sélectionnez une paire de clés qui vous permet de vous connecter en toute sécurité à l'instance du moteur NetApp GenAI.

7. Compléter la section **ai Engine** :

- Nom de l'instance** : si vous le souhaitez, sélectionnez **Définir le nom de l'instance** et saisissez un nom personnalisé pour l'instance du moteur d'IA. Le nom de l'instance apparaît dans la console de gestion AWS et dans la zone d'informations sur l'infrastructure de Workload Factory et peut vous aider à suivre les ressources de Workload Factory.

8. Sélectionnez **déployer** pour commencer le déploiement.



Si le déploiement échoue avec une erreur d'informations d'identification, vous pouvez obtenir plus de détails sur l'erreur en sélectionnant les liens hypertexte dans le message d'erreur. Vous pouvez voir une liste d'autorisations manquantes ou bloquées, ainsi qu'une liste d'autorisations dont la charge de travail GenAI a besoin pour pouvoir déployer l'infrastructure GenAI.

## Résultat

Workload Factory commence à déployer l'infrastructure du chatbot. Ce processus peut prendre jusqu'à 10 minutes.

Au cours du processus de déploiement, les éléments suivants sont configurés :

- Le réseau est configuré avec les terminaux privés.
- Le rôle IAM, le profil d'instance et le groupe de sécurité sont créés.
- L'instance de machine virtuelle du moteur GenAI est déployée.
- Amazon Bedrock est configuré pour envoyer des journaux à Amazon CloudWatch, en utilisant un groupe de journaux avec le préfixe /aws/bedrock/.
- Le moteur GenAI est configuré pour envoyer des journaux à Amazon CloudWatch Logs, à l'aide d'un groupe de journaux portant le nom /netapp/wlmai/<tenancyAccountId>/randomId , où <tenancyAccountId> est le "["ID de compte de la console NetApp"](#) pour l'utilisateur actuel.

## **Informations sur le copyright**

Copyright © 2025 NetApp, Inc. Tous droits réservés. Imprimé aux États-Unis. Aucune partie de ce document protégé par copyright ne peut être reproduite sous quelque forme que ce soit ou selon quelque méthode que ce soit (graphique, électronique ou mécanique, notamment par photocopie, enregistrement ou stockage dans un système de récupération électronique) sans l'autorisation écrite préalable du détenteur du droit de copyright.

Les logiciels dérivés des éléments NetApp protégés par copyright sont soumis à la licence et à l'avis de non-responsabilité suivants :

CE LOGICIEL EST FOURNI PAR NETAPP « EN L'ÉTAT » ET SANS GARANTIES EXPRESSES OU TACITES, Y COMPRIS LES GARANTIES TACITES DE QUALITÉ MARCHANDE ET D'ADÉQUATION À UN USAGE PARTICULIER, QUI SONT EXCLUES PAR LES PRÉSENTES. EN AUCUN CAS NETAPP NE SERA TENU POUR RESPONSABLE DE DOMMAGES DIRECTS, INDIRECTS, ACCESSOIRES, PARTICULIERS OU EXEMPLAIRES (Y COMPRIS L'ACHAT DE BIENS ET DE SERVICES DE SUBSTITUTION, LA PERTE DE JOUSSANCE, DE DONNÉES OU DE PROFITS, OU L'INTERRUPTION D'ACTIVITÉ), QUELLES QU'EN SOIENT LA CAUSE ET LA DOCTRINE DE RESPONSABILITÉ, QU'IL S'AGISSE DE RESPONSABILITÉ CONTRACTUELLE, STRICTE OU DÉLICTUELLE (Y COMPRIS LA NÉGLIGENCE OU AUTRE) DÉCOULANT DE L'UTILISATION DE CE LOGICIEL, MÊME SI LA SOCIÉTÉ A ÉTÉ INFORMÉE DE LA POSSIBILITÉ DE TELS DOMMAGES.

NetApp se réserve le droit de modifier les produits décrits dans le présent document à tout moment et sans préavis. NetApp décline toute responsabilité découlant de l'utilisation des produits décrits dans le présent document, sauf accord explicite écrit de NetApp. L'utilisation ou l'achat de ce produit ne concède pas de licence dans le cadre de droits de brevet, de droits de marque commerciale ou de tout autre droit de propriété intellectuelle de NetApp.

Le produit décrit dans ce manuel peut être protégé par un ou plusieurs brevets américains, étrangers ou par une demande en attente.

LÉGENDE DE RESTRICTION DES DROITS : L'utilisation, la duplication ou la divulgation par le gouvernement sont sujettes aux restrictions énoncées dans le sous-paragraphe (b)(3) de la clause Rights in Technical Data-Noncommercial Items du DFARS 252.227-7013 (février 2014) et du FAR 52.227-19 (décembre 2007).

Les données contenues dans les présentes se rapportent à un produit et/ou service commercial (tel que défini par la clause FAR 2.101). Il s'agit de données propriétaires de NetApp, Inc. Toutes les données techniques et tous les logiciels fournis par NetApp en vertu du présent Accord sont à caractère commercial et ont été exclusivement développés à l'aide de fonds privés. Le gouvernement des États-Unis dispose d'une licence limitée irrévocable, non exclusive, non cessible, non transférable et mondiale. Cette licence lui permet d'utiliser uniquement les données relatives au contrat du gouvernement des États-Unis d'après lequel les données lui ont été fournies ou celles qui sont nécessaires à son exécution. Sauf dispositions contraires énoncées dans les présentes, l'utilisation, la divulgation, la reproduction, la modification, l'exécution, l'affichage des données sont interdits sans avoir obtenu le consentement écrit préalable de NetApp, Inc. Les droits de licences du Département de la Défense du gouvernement des États-Unis se limitent aux droits identifiés par la clause 252.227-7015(b) du DFARS (février 2014).

## **Informations sur les marques commerciales**

NETAPP, le logo NETAPP et les marques citées sur le site <http://www.netapp.com/TM> sont des marques déposées ou des marques commerciales de NetApp, Inc. Les autres noms de marques et de produits sont des marques commerciales de leurs propriétaires respectifs.