Contrôles de sécurité de l'IA générative

Information Security for Workload Factory

NetApp
September 16, 2026

Sommaire

Contrôles de sécurité de l'IA générative

Découvrez les contrôles d'IA dans NetApp Workload Factory

NetApp Workload Factory intègre des fonctionnalités d'IA générative qui accélèrent le diagnostic tout en appliquant des contrôles de sécurité pour l'accès au modèle, la portée des données et le comportement d'exécution.

Workload Factory active par défaut les diagnostics par IA. Vous pouvez désactiver les fonctionnalités d'IA générative à tout moment via les paramètres **Administration** de Workload Factory. "En savoir plus sur la gestion des fonctionnalités GenAI."

Aperçu des fonctionnalités d'IA à portée de sécurité

Workload Factory applique actuellement l'IA générative aux fonctionnalités suivantes :

- Assistant Ask Me (RAG avec documentation NetApp sélectionnée, réponses étayées par des sources)
- Analyse de la latence
- analyse des événements EMS
- Analyse du journal des erreurs

Présentation d'AWS Bedrock pour les diagnostics d'IA

Workload Factory utilise Amazon Bedrock pour l'inférence IA. L'inférence s'exécute dans votre compte AWS, avec vos identifiants et votre profil d'inférence configurés.

Informations connexes

- "Activation d'Amazon Bedrock pour les diagnostics IA de NetApp Workload Factory"
- "Limites des outils d'IA pour NetApp Workload Factory"
- "Paramètres du modèle d'IA et facturation pour NetApp Workload Factory"

Activation d'Amazon Bedrock pour les diagnostics IA de NetApp Workload Factory

NetApp Workload Factory diagnostics IA utilisent Amazon Bedrock pour analyser les événements, et la fonctionnalité est activée par défaut. Avant que Workload Factory puisse invoquer le modèle Bedrock sélectionné et collecter les données de diagnostic, vous devez configurer les autorisations IAM, un profil d'inférence et une liaison Lambda connectée avec la capacité SSH. Si l'un de ces composants est manquant ou si la configuration Bedrock est supprimée, les diagnostics IA deviennent indisponibles.

Les diagnostics par IA sont activés par défaut.

Avant de commencer

- Assurez-vous que votre compte inclut un profil d'inférence AWS Bedrock.
- Disposer d'une liaison Lambda connectée avec une capacité SSH pour la collecte de données de diagnostic.

Étapes

1. Accordez les autorisations Bedrock dans votre rôle IAM :
    - `bedrock:InvokeModel`
    - `bedrock:InvokeModelWithResponseStream`
    - `bedrock:ListInferenceProfiles`
    - `bedrock:GetInferenceProfile`
2. Configurez l'ARN du profil d'inférence dans les paramètres de Workload Factory.

Si l'un des prérequis est manquant, le système signale le composant manquant spécifique et les fonctionnalités d'IA restent inactives.

Désactiver les diagnostics IA

Pour désactiver les diagnostics IA :

- Supprimez les autorisations Bedrock de votre rôle IAM, ou
- Supprimez la configuration Bedrock dans les paramètres de Workload Factory.

Les diagnostics IA deviennent immédiatement indisponibles et aucun traitement des données ne se produit.

Profils d'inférence régionaux et interrégionaux

L'inférence Bedrock s'exécute dans la région AWS spécifiée par votre profil d'inférence. Si vous utilisez un profil d'inférence interrégional, AWS peut acheminer la requête vers n'importe quelle région dans la configuration du profil, conformément au comportement standard d'AWS Bedrock. Les données ne quittent pas l'infrastructure AWS.

Limites des outils d'IA pour NetApp Workload Factory

NetApp Workload Factory impose des limites strictes à la manière dont ses fonctionnalités d'IA interagissent avec vos environnements AWS et ONTAP, afin que vous puissiez faire confiance aux diagnostics assistés par l'IA sans risquer de modifications involontaires. Amazon Bedrock utilise des outils de diagnostic AWS CLI et ONTAP CLI en lecture seule qui bloquent les commandes de modification et limitent la sortie par taille et par exécution, et toutes les données utilisées sont uniquement transitoires et ne sont conservées ni par NetApp Workload Factory ni par AWS.

Contrôles de sécurité des outils

Outil	Capacité	Contrôles de sécurité
AWS CLI (lecture seule)	Exécutez les commandes AWS CLI en lecture seule (describe, list, get)	Bloque tous les verbes de modification (créer, supprimer, modifier, mettre à jour, ajouter, retirer). Maximum 10 commandes par étape. Sortie tronquée à 20 Ko.
Interface de ligne de commande ONTAP (lecture seule)	Exécutez des commandes ONTAP CLI en lecture seule via SSH	Bloque tous les verbes de modification (supprimer, détruire, créer, modifier, déplacer). Sortie tronquée.

L'agent ne peut ni modifier, ni créer, ni supprimer de ressources. Il peut uniquement observer et diagnostiquer.

Limites de la conservation des données et de l'entraînement du modèle

Conformément à la politique d'AWS, Amazon Bedrock ne stocke ni n'utilise vos entrées et sorties pour entraîner ou améliorer les modèles de base. Pour l'inférence à la demande (utilisée par Workload Factory), AWS traite vos données de manière temporaire et ne les conserve pas après avoir renvoyé la réponse.

Paramètres du modèle d'IA et facturation pour NetApp Workload Factory

NetApp Workload Factory définit la configuration du modèle et les limites d'utilisation et de facturation pertinentes pour la gouvernance de l'IA. Le modèle configuré utilise des paramètres déterministes et renvoie du JSON structuré pour garantir la cohérence des résultats d'analyse. AWS facture l'inférence Amazon Bedrock sur votre compte, tandis que Workload Factory suit l'utilisation de l'IA et fournit une visibilité mensuelle des coûts par compte. Ces sections décrivent les paramètres du modèle et les limites applicables à la consommation d'IA.

Modèle et paramètres

Paramètre	Valeur
Modèle	Anthropic Claude (utilisant un profil d'inférence interrégional)
Température	0 (déterministe, sans aléa)
Jetons de sortie maximum	2 048
Nombre maximal d'étapes de raisonnement	15 par analyse
Format de sortie	JSON structuré (résumé, gravité, cause racine, actions correctives)

Limites de facturation et d'utilisation

- AWS facture l'inférence Bedrock à votre compte (votre profil d'inférence, vos identifiants).
- Workload Factory assure le suivi interne de l'utilisation de l'IA.
- Workload Factory offre une visibilité mensuelle des coûts par compte.

"Suivez les coûts de stockage dans NetApp Workload Factory"

Assistant IA Ask Me

Architecture de sécurité Ask Me pour NetApp Workload Factory

Ask Me est un assistant génératif basé sur l'IA, intégré à NetApp Workload Factory. Il offre une assistance conversationnelle en temps réel pour Amazon FSx for NetApp ONTAP et les sujets Workload Factory. Les réponses s'appuient sur une documentation NetApp vérifiée, et Ask Me ne peut pas accéder à Internet ni utiliser les données client pour entraîner le modèle. Plusieurs couches de sécurité bloquent les requêtes malveillantes, limitent les réponses aux domaines pris en charge et empêchent que les saisies utilisateur ne remplacent les instructions système. Lorsque Ask Me utilise des outils, les limites d'autorisation, l'application de la lecture seule pour les requêtes, un environnement de test éphémère et la journalisation des audits permettent de restreindre et de surveiller l'exécution.

Vous pouvez désactiver Ask Me à tout moment via les paramètres **Administration** de Workload Factory, ce qui désactive l'assistant pour votre compte utilisateur. "En savoir plus sur la gestion des fonctionnalités GenAI."

Architecture RAG dans Ask Me

Ask Me utilise la génération augmentée par récupération (RAG).

- Base de connaissances organisée : les réponses s'appuient sur un corpus géré de documentation NetApp vérifiée et publiée.
- Évaluation et filtrage de la recherche : seul le contenu suffisamment pertinent est inclus dans le contexte du modèle.
- Aucun accès à Internet : le modèle ne peut pas récupérer, parcourir ou extraire de contenu externe.
- Attribution des sources : les réponses incluent des références aux documents sources utilisés.

Sécurité multicouche des invites

- Couche 1 : Le classificateur de sécurité (LLM-as-a-judge) bloque les requêtes malveillantes (injection de prompt, élévation de privilèges, exfiltration de données, ingénierie sociale, violations de politiques).

  Le système enregistre les requêtes bloquées, et le modèle génératif ne les voit jamais.

- Couche 2 : Le renforcement des invites système empêche les entrées de l'utilisateur de prendre le pas sur les instructions du système.
- Couche 3 : La portée du domaine limite les réponses aux sujets FSx ONTAP et NetApp Workload Factory.
- Couche 4 : La gouvernance de l'utilisation des outils valide les paramètres et exécute les requêtes dans un environnement sécurisé ; le modèle ne peut pas exécuter d'opérations arbitraires.

Isolation et confidentialité des données du modèle

- Aucun entraînement du modèle sur les données client
- Isolation au niveau des requêtes (aucune fuite de données entre locataires)

- Contexte de conversation à l’échelle de la session avec un historique limité
- Aucune mémoire de modèle persistante
- Infrastructure d’inférence gérée

Utilisation d’outils et sécurité agentique

- Outils à autorisation limitée (limites des permissions utilisateur appliquées)
- Délais d’exécution stricts des outils
- Bac à sable éphémère, limité à la session, pour les données récupérées
- Application des requêtes en lecture seule avec liste d’autorisation
- Auditabilité des appels d’outils avec nettoyage des paramètres sensibles

Contrôle d’accès Ask Me pour NetApp Workload Factory

Dans NetApp Workload Factory, la sécurité d’accès Ask Me repose sur des sessions authentifiées, la responsabilité au niveau de l’utilisateur et la gestion protégée des secrets lors des opérations d’exécution. Les contrôles d’authentification valident chaque session et associent les interactions à un utilisateur spécifique. Les informations d’identification sensibles sont récupérées depuis un coffre-fort de secrets géré lors de l’exécution et supprimées des journaux. Le service applique également des contrôles d’infrastructure, notamment l’exécution de conteneurs sans privilèges root et une liste d’autorisation CORS restreinte.

Authentification

- Authentification JWT signée avec validation cryptographique (y compris les vérifications de l’audience, de l’émetteur et de l’algorithme telles que RS256)
- Authentification appliquée par le middleware à la frontière du service
- Définition du périmètre d’identité de l’utilisateur afin que les interactions soient attribuables à un utilisateur spécifique

Protection des identifiants et gestion des secrets

- Stockage sécurisé des identifiants : les identifiants sensibles utilisés par le service sont stockés dans un coffre-fort de secrets géré et récupérés lors de l’exécution. Aucun secret n’est stocké dans le code de l’application, les fichiers de configuration ou les images de conteneur.
- Nettoyage des journaux : les valeurs sensibles (identifiants, jetons, mots de passe, clés d’accès, certificats) sont supprimées des journaux avant leur enregistrement.

Sécurité des infrastructures

- Exécution de conteneur non root
- CORS restreint à une liste blanche explicite de domaines de confiance NetApp

Modes d’exécution Ask Me pour NetApp Workload Factory

NetApp Workload Factory propose plusieurs modes d’exécution pour Ask Me, chacun

présentant des caractéristiques de sécurité et de confidentialité différentes. Lorsqu'Ask Me utilise des intégrations d'outils, par exemple pour interroger les ressources Workload Factory du client, l'exécution des outils est contrôlée par des dispositifs de protection multicouches. L'exécution avec outils s'effectue dans le cadre des autorisations de l'utilisateur authentifié et applique des limites à la durée et à la portée de chaque opération. Les données récupérées restent dans un environnement de session éphémère qui bloque tout accès externe et sont détruites à la fin de la session. Des contrôles indépendants des invites et du code imposent des requêtes en lecture seule, tandis que les journaux d'audit enregistrent l'activité des outils après suppression des valeurs sensibles.

Garde-fous pour l'exécution des outils

- Les outils à portée d'autorisation fonctionnent dans le cadre des autorisations de l'utilisateur authentifié.
- Le délai d'expiration de l'exécution de l'outil limite les opérations de longue durée.
- Le bac à sable de données éphémères stocke les données récupérées par l'outil dans le cadre de la session, bloque l'accès externe, les E/S de fichiers, les appels réseau et le chargement d'extensions au niveau du moteur, et est détruit à la fin de la session.
- L'application du contrôle des requêtes en lecture seule valide les requêtes générées via une liste d'autorisation stricte.
- La sécurité appliquée par invite et la sécurité appliquée par code sont appliquées indépendamment.
- L'audit des appels d'outils enregistre le nom de l'outil, ses paramètres, les horodatages et l'état du résultat, avec un filtrage des valeurs sensibles.

Contrôles de sortie

- Application des limites de jetons
- Sortie déterministe pour les opérations critiques de classification/sécurité (température 0)
- Diffusion en continu avec arrêt anticipé et nettoyage

Demandez-moi la confidentialité et la disponibilité de NetApp Workload Factory

Dans NetApp Workload Factory, les contrôles de confidentialité Ask Me limitent l'exposition des données, maintiennent l'isolation du traitement des sessions et préservent les frontières de confidentialité pour les interactions utilisateur. Vos entrées sont traitées par un service d'IA géré qui ne les utilise pas pour entraîner ou améliorer les modèles de base. Les données opérationnelles récupérées restent dans un stockage éphémère en mémoire pendant la durée de la session et ne sont ni conservées ni partagées. Une chaîne de repli multi-modèles répartie sur plusieurs régions cloud garantit la disponibilité lorsque le modèle principal est limité ou indisponible, tandis que les mêmes contrôles de sécurité s'appliquent à tous les modèles de repli.

Gestion des données et confidentialité

- Données utilisateur dans les invites : traitées par un service d'IA géré qui n'utilise pas vos entrées pour entraîner ou améliorer les modèles de base.

- Contenu de la base de connaissances : documentation NetApp sélectionnée et publiée uniquement ; vos données ne sont pas incluses.
- Données opérationnelles : récupérées uniquement lorsqu'un utilisateur interroge ses propres ressources ; stockées de manière éphémère en mémoire pendant la durée de la session et non conservées ni partagées.
- Journalisation des audits : les métadonnées des requêtes (identifiant utilisateur anonymisé, horodatages, durées) sont enregistrées, avec les champs sensibles supprimés.
- Données de retour d'information : stockées séparément et liées à un identifiant de requête, non à des informations permettant d'identifier une personne au-delà de l'identifiant utilisateur anonymisé.

Contrôles de disponibilité et d'isolation

Ask Me utilise une chaîne de repli multi-modèles répartie sur plusieurs régions cloud.

Si le modèle principal est limité ou indisponible, le système peut basculer vers des modèles alternatifs.

Tous les modèles sont soumis aux mêmes contrôles de sécurité, au même renforcement du système de sécurité et aux mêmes garanties d'isolation.

Informations sur le copyright

Informations sur les marques commerciales