Controlli di sicurezza per l'intelligenza artificiale generativa

Information Security for Workload Factory

NetApp
September 16, 2026

Sommario

Controlli di sicurezza per l'intelligenza artificiale generativa

Scopri i controlli AI in NetApp Workload Factory

NetApp Workload Factory include funzionalità di intelligenza artificiale generativa che accelerano la diagnosi mentre applicano controlli di sicurezza per l'accesso al modello, l'ambito dei dati e il comportamento in fase di esecuzione.

Workload Factory abilita la diagnostica basata sull'AI per impostazione predefinita. Puoi disabilitare le funzionalità di AI generativa in qualsiasi momento tramite le impostazioni di **Administration** di Workload Factory. "Scopri di più sulla gestione delle funzionalità di GenAI."

Panoramica delle funzionalità di IA con ambito di sicurezza

Attualmente Workload Factory applica l'intelligenza artificiale generativa alle seguenti funzionalità:

- Assistente Chiedimi (RAG con documentazione NetApp curata, risposte supportate da fonti)
- Analisi della latenza
- analisi degli eventi EMS
- Analisi del registro degli errori

Panoramica di AWS Bedrock per la diagnostica dell'AI

Workload Factory utilizza Amazon Bedrock per l'inferenza AI. L'inferenza viene eseguita nel tuo account AWS, con le credenziali e il profilo di inferenza configurati.

Informazioni correlate

- "Adesione facoltativa ad Amazon Bedrock per la diagnostica AI di NetApp Workload Factory"
- "Limiti degli strumenti di intelligenza artificiale per NetApp Workload Factory"
- "Parametri del modello AI e fatturazione per NetApp Workload Factory"

Adesione facoltativa ad Amazon Bedrock per la diagnostica AI di NetApp Workload Factory

NetApp Workload Factory AI diagnostics utilizza Amazon Bedrock per analizzare gli eventi ed è abilitata per impostazione predefinita. Prima che Workload Factory possa richiamare il modello Bedrock selezionato e raccogliere i dati diagnostici, devi configurare le autorizzazioni IAM, un profilo di inferenza e un collegamento Lambda con funzionalità SSH. Se uno di questi componenti manca o la configurazione Bedrock viene rimossa, la diagnostica AI non sarà disponibile.

La diagnostica AI è abilitata per impostazione predefinita.

Prima di iniziare

- Assicurati che il tuo account includa un profilo di inferenza AWS Bedrock.
- Hai un collegamento Lambda connesso con funzionalità SSH per la raccolta di dati diagnostici.

Passaggi

1. Concedi le autorizzazioni Bedrock nel tuo ruolo IAM:
    - `bedrock:InvokeModel`
    - `bedrock:InvokeModelWithResponseStream`
    - `bedrock:ListInferenceProfiles`
    - `bedrock:GetInferenceProfile`
2. Configura l'ARN del profilo di inferenza nelle impostazioni di Workload Factory.

Se manca uno qualsiasi dei prerequisiti, il sistema segnala il componente specifico mancante e le funzionalità AI rimangono inattive.

Disabilita la diagnostica dell'IA

Per disattivare la diagnostica AI:

- Rimuovi le autorizzazioni Bedrock dal tuo ruolo IAM, oppure
- Elimina la configurazione Bedrock nelle impostazioni di Workload Factory.

La diagnostica basata sull'AI diventa immediatamente non disponibile e non si verifica alcuna elaborazione residua dei dati.

Profili di inferenza regionali e tra regioni

L'inferenza Bedrock viene eseguita nella regione AWS specificata dal tuo profilo di inferenza. Se usi un profilo di inferenza interregionale, AWS potrebbe instradare la richiesta a qualsiasi regione nella configurazione del profilo, secondo il comportamento standard di AWS Bedrock. I dati non escono dall'infrastruttura AWS.

Limiti degli strumenti di intelligenza artificiale per NetApp Workload Factory

NetApp Workload Factory impone rigidi limiti su come le sue funzionalità di intelligenza artificiale interagiscono con i tuoi ambienti AWS e ONTAP, così puoi affidarti alla diagnostica assistita dall'IA senza rischiare modifiche indesiderate. Amazon Bedrock utilizza strumenti di diagnostica AWS CLI e ONTAP CLI in sola lettura che bloccano i comandi di modifica e limitano l'output per dimensione ed esecuzione, e tutti i dati utilizzati sono solo temporanei e non vengono conservati da NetApp Workload Factory o AWS.

Controlli di sicurezza degli strumenti

Strumento	Capacità	Controlli di sicurezza
AWS CLI (sola lettura)	Esegui comandi AWS CLI di sola lettura (describe, list, get)	Blocca tutti i verbi di modifica (create, delete, modify, update, put, remove). Massimo 10 comandi per step. Output troncato a 20 KB.
ONTAP CLI (sola lettura)	Esegui comandi CLI ONTAP in sola lettura tramite SSH	Blocca tutti i verbi di modifica (elimina, distruggi, crea, modifica, sposta). Output troncato.

L'agente non può modificare, creare o eliminare risorse. Può solo osservare e diagnosticare.

Limiti di conservazione dei dati e di addestramento del modello

In conformità con le policy AWS, Amazon Bedrock non memorizza né utilizza i tuoi input e output per addestrare o migliorare i foundation model. Per l'inferenza on-demand (usata da Workload Factory), AWS elabora i tuoi dati in modo transitorio e non li conserva dopo aver restituito la risposta.

Parametri del modello AI e fatturazione per NetApp Workload Factory

NetApp Workload Factory definisce la configurazione del modello e i limiti di utilizzo e fatturazione rilevanti per la governance dell'AI. Il modello configurato utilizza parametri deterministici e restituisce JSON strutturato per garantire risultati di analisi coerenti. AWS fattura l'inferenza di Amazon Bedrock al tuo account, mentre Workload Factory tiene traccia dell'utilizzo dell'AI e fornisce visibilità sui costi mensili per account. Queste sezioni descrivono i parametri del modello e i limiti che si applicano al consumo di AI.

Modello e parametri

Parametro	Valore
Modello	Anthropic Claude (utilizzando un profilo di inferenza interregionale)
Temperatura	0 (deterministico, senza casualità)
Token di output massimi	2.048
Numero massimo di passaggi di ragionamento	15 per analisi
Formato di output	JSON strutturato (riepilogo, gravità, causa principale, azioni correttive)

Limiti di fatturazione e utilizzo

- AWS addebita l'inferenza Bedrock al tuo account (il tuo profilo di inferenza, le tue credenziali).
- Workload Factory tiene traccia internamente dell'utilizzo dell'AI.
- Workload Factory offre visibilità sui costi mensili per account.

  "Monitora i costi di archiviazione in NetApp Workload Factory"

Chiedimi assistente AI

Chiedimi l’architettura di sicurezza per NetApp Workload Factory

Ask Me è un assistente generativo basato sull’intelligenza artificiale integrato in NetApp Workload Factory. Fornisce supporto conversazionale in real-time per Amazon FSx for NetApp ONTAP e argomenti di Workload Factory. Le risposte si basano su documentazione NetApp verificata e Ask Me non può accedere a Internet pubblico né utilizzare i dati dei clienti per addestrare il modello. Diversi livelli di sicurezza aiutano a bloccare le query dannose, limitare le risposte ai domini supportati e impedire che l’input dell’utente sovrascriva le istruzioni di sistema. Quando Ask Me utilizza strumenti, i limiti di autorizzazione, l’applicazione di query di sola lettura, una sandbox temporanea e la registrazione degli eventi aiutano a limitare e monitorare l’esecuzione.

Puoi disattivare Ask Me in qualsiasi momento tramite le impostazioni di **Amministrazione** di Workload Factory, il che disabilita l’assistente per il tuo account utente. "Scopri di più sulla gestione delle funzionalità di GenAI."

Architettura RAG in Ask Me

Ask Me utilizza la generazione aumentata tramite recupero (RAG).

- Base di conoscenza curata: le risposte si basano su un corpus gestito di documentazione NetApp verificata e pubblicata.
- Valutazione e filtraggio del recupero: solo i contenuti sufficientemente pertinenti sono inclusi nel contesto del modello.
- Nessun accesso a Internet: il modello non può recuperare, visualizzare o estrarre contenuti esterni.
- Attribuzione della fonte: le risposte includono riferimenti ai documenti di origine utilizzati.

Sicurezza multi-livello dei prompt

- Livello 1: il classificatore di sicurezza (LLM-as-a-judge) blocca le query dannose (iniezione di prompt, escalation dei privilegi, esfiltrazione di dati, ingegneria sociale, violazioni delle policy).

  Il sistema registra le query bloccate e il modello generativo non le vede mai.

- Livello 2: L’irrobustimento del prompt di sistema impedisce che l’input dell’utente sovrascriva le istruzioni di sistema.
- Livello 3: l’ambito del dominio limita le risposte agli argomenti FSx ONTAP e Workload Factory.
- Livello 4: La governance dell’utilizzo degli strumenti convalida i parametri ed esegue le query in un ambiente protetto; il modello non può eseguire operazioni arbitrarie.

Isolamento e privacy dei dati del modello

- Nessun addestramento del modello sui dati dei clienti
- Isolamento a livello di richiesta (nessuna perdita di dati tra tenant)
- Contesto di conversazione a livello di sessione con cronologia delimitata
- Nessuna memoria modello persistente

- Infrastruttura di inferenza gestita

Uso degli strumenti e sicurezza agentica

- Strumenti con ambito di autorizzazione (limiti di autorizzazione utente applicati)
- Rigorosi timeout di esecuzione degli strumenti
- Sandbox effimera, limitata alla sessione, per i dati recuperati
- Applicazione delle query di sola lettura con allowlisting
- Tracciabilità delle chiamate degli strumenti con pulizia dei parametri sensibili

Chiedimi il controllo degli accessi per NetApp Workload Factory

In NetApp Workload Factory, la sicurezza di accesso di Ask Me si basa su sessioni autenticate, responsabilità a livello utente e gestione protetta dei segreti durante le operazioni in fase di esecuzione. I controlli di autenticazione convalidano ogni sessione e associano le interazioni a un utente specifico. Le credenziali sensibili vengono recuperate da un archivio di segreti gestito in fase di esecuzione e oscurate nei log. Il servizio applica anche controlli infrastrutturali, tra cui l'esecuzione di container senza privilegi di root e una lista di autorizzazioni CORS limitata.

Autenticazione

- Autenticazione JWT firmata con convalida crittografica (inclusi controlli su audience, issuer e algoritmo come RS256)
- Autenticazione imposta dal middleware al confine del servizio
- Definizione dell'ambito dell'identità utente così che le interazioni siano attribuibili a un utente specifico

Protezione delle credenziali e gestione dei segreti

- Archiviazione sicura nel vault: le credenziali sensibili utilizzate dal servizio vengono archiviate in un vault di segreti gestito e recuperate in fase di esecuzione. Nessun segreto viene memorizzato nel codice dell'applicazione, nei file di configurazione o nelle immagini dei container.
- Pulizia dei log: i valori sensibili (credenziali, token, password, chiavi di accesso, certificati) vengono oscurati dai log prima che siano scritti.

Sicurezza delle infrastrutture

- Esecuzione di container non root
- CORS limitato a una lista esplicita di domini attendibili NetApp

Chiedimi le modalità di esecuzione per NetApp Workload Factory

NetApp Workload Factory offre diverse modalità di esecuzione per Ask Me, ognuna con caratteristiche di sicurezza e privacy differenti. Quando Ask Me utilizza integrazioni con strumenti, ad esempio interrogando le risorse Workload Factory del cliente, l'esecuzione dello strumento è controllata tramite misure di sicurezza a più livelli. L'esecuzione abilitata dallo strumento opera entro i limiti delle autorizzazioni dell'utente autenticato e

applica limiti alla durata e all'ambito di ciascuna operazione. I dati recuperati rimangono in una sandbox di sessione temporanea che blocca l'accesso esterno e viene eliminata al termine della sessione. Controlli indipendenti su prompt e codice impongono query di sola lettura, mentre i registri di audit registrano l'attività dello strumento con i valori sensibili rimossi.

Salvaguardie per l'esecuzione degli strumenti

- Gli strumenti con ambito di autorizzazione operano entro le autorizzazioni dell'utente autenticato.
- Il timeout di esecuzione dello strumento limita le operazioni di lunga durata.
- La sandbox di dati effimeri memorizza i dati recuperati dagli strumenti nell'ambito della sessione, blocca l'accesso esterno, l'I/O dei file, le chiamate di rete e il caricamento delle estensioni a livello del motore e viene distrutta al termine della sessione.
- L'applicazione della modalità di sola lettura alle query convalida le query generate tramite una rigorosa allowlist.
- La sicurezza basata su prompt e quella basata su codice vengono applicate in modo indipendente.
- La tracciabilità delle chiamate agli strumenti registra il nome dello strumento, i parametri, i timestamp e lo stato del risultato, con la rimozione dei valori sensibili.

controlli di output

- applicazione del limite di token
- Output deterministico per operazioni di classificazione/sicurezza-critical (temperatura 0)
- Streaming con terminazione anticipata e pulizia

Chiedimi informazioni su privacy e disponibilità per NetApp Workload Factory

In NetApp Workload Factory, i controlli sulla privacy di Ask Me limitano l'esposizione dei dati, mantengono isolata l'elaborazione delle sessioni e preservano i confini della privacy per le interazioni con gli utenti. I tuoi input vengono elaborati da un servizio di intelligenza artificiale gestito che non li utilizza per addestrare o migliorare i modelli di base. I dati operativi recuperati rimangono in una memoria temporanea, in-memory, per la durata della sessione e non vengono memorizzati o condivisi. Una catena di fallback multi-modello distribuita su più regioni cloud garantisce la disponibilità quando il modello principale è limitato o non disponibile, mentre gli stessi controlli di sicurezza si applicano ai modelli di fallback.

Gestione dei dati e privacy

- Dati utente nei prompt: elaborati da un servizio AI gestito che non utilizza i tuoi input per addestrare o migliorare i foundation models.
- Contenuto della knowledge base: solo documentazione curata e pubblicata NetApp; i tuoi dati non sono inclusi.
- Dati operativi: recuperati solo quando un utente interroga le proprie risorse; conservati in memoria temporanea per la durata della sessione e non persistenti né condivisi.
- Registrazione degli eventi di controllo: metadati delle query (ID utente anonimizzato, timestamp, durate)

registrati, con i campi sensibili oscurati.

- Dati di feedback: memorizzati separatamente e collegati a un ID di query, non a informazioni di identificazione personale oltre all'ID utente anonimizzato.

Controlli di disponibilità e isolamento

Ask Me utilizza una catena di fallback multi-modello su più regioni cloud.

Se il modello principale è sovraccarico o non disponibile, il sistema può passare a modelli alternativi.

Tutti i modelli sono soggetti agli stessi controlli di sicurezza, allo stesso hardening del sistema e alle stesse garanzie di isolamento.