Controles de seguridad para la IA generativa

Information Security for Workload Factory

NetApp
September 16, 2026

Tabla de contenidos

Controles de seguridad para la IA generativa . . . 1
Conoce los controles de IA en NetApp Workload Factory . . . 1
Descripción general de las funciones de IA centradas en la seguridad . . . 1
Descripción general de AWS Bedrock para diagnósticos de IA . . . 1
Información relacionada . . . 1
Habilitación de Amazon Bedrock para los diagnósticos de IA de NetApp Workload Factory . . . 1
Antes de empezar . . . 2
Pasos . . . 2
Desactivar el diagnóstico de IA . . . 2
Perfiles de inferencia regionales e interregionales . . . 2
Límites de las herramientas de IA para NetApp Workload Factory . . . 2
Controles de seguridad de las herramientas . . . 3
Límites de la conservación de datos y del entrenamiento de modelos . . . 3
Parámetros del modelo de IA y facturación para NetApp Workload Factory . . . 3
Modelo y parámetros . . . 3
Límites de facturación y uso . . . 3
Asistente de IA Ask Me . . . 4
Arquitectura de seguridad de Ask Me para NetApp Workload Factory . . . 4
Control de acceso de Ask Me para NetApp Workload Factory . . . 5
Modos de ejecución de Ask Me para NetApp Workload Factory . . . 6
Consúltame sobre la privacidad y la disponibilidad de NetApp Workload Factory . . . 6

Controles de seguridad para la IA generativa

Conoce los controles de IA en NetApp Workload Factory

NetApp Workload Factory incluye funciones de IA generativa que aceleran el diagnóstico, mientras aplican controles de seguridad para el acceso a modelos, el alcance de los datos y el comportamiento en tiempo de ejecución.

Workload Factory habilita el diagnóstico mediante IA de forma predeterminada. Puedes desactivar las funciones de IA generativa en cualquier momento a través de la configuración de **Administración** de Workload Factory. "Obtén más información sobre cómo gestionar las funciones de GenAI."

Descripción general de las funciones de IA centradas en la seguridad

Workload Factory aplica actualmente la IA generativa a las siguientes funciones:

- Asistente «Ask Me» (RAG con documentación seleccionada de NetApp y respuestas respaldadas por fuentes)
- Análisis de latencia
- Análisis de eventos de EMS
- Análisis del registro de errores

Descripción general de AWS Bedrock para diagnósticos de IA

Workload Factory utiliza Amazon Bedrock para la inferencia de IA. La inferencia se ejecuta en tu cuenta de AWS, con las credenciales y el perfil de inferencia que hayas configurado.

Información relacionada

- "Habilitación de Amazon Bedrock para los diagnósticos de IA de NetApp Workload Factory"
- "Límites de las herramientas de IA para NetApp Workload Factory"
- "Parámetros del modelo de IA y facturación para NetApp Workload Factory"

Habilitación de Amazon Bedrock para los diagnósticos de IA de NetApp Workload Factory

Los diagnósticos de IA de NetApp Workload Factory utilizan Amazon Bedrock para analizar eventos, y esta función está habilitada de forma predeterminada. Para que Workload Factory pueda invocar el modelo de Bedrock seleccionado y recopilar datos de diagnóstico, debes configurar los permisos de IAM, un perfil de inferencia y un enlace Lambda conectado con capacidad SSH. Si falta alguno de estos componentes o se elimina la configuración de Bedrock, los diagnósticos de IA dejarán de estar disponibles.

Los diagnósticos de IA están activados de forma predeterminada.

Antes de empezar

- Asegúrate de que tu cuenta incluya un perfil de inferencia de AWS Bedrock.
- Tener un enlace Lambda conectado con capacidad SSH para la recopilación de datos de diagnóstico.

Pasos

1. Otorga permisos de Bedrock a tu rol de IAM:
    - `bedrock:InvokeModel`
    - `bedrock:InvokeModelWithResponseStream`
    - `bedrock:ListInferenceProfiles`
    - `bedrock:GetInferenceProfile`
2. Configura el ARN del perfil de inferencia en los ajustes de Workload Factory.

Si falta algún requisito previo, el sistema indica qué componente concreto falta y las funciones de IA permanecen inactivas.

Desactivar el diagnóstico de IA

Para desactivar los diagnósticos de IA:

- Elimina los permisos de Bedrock de tu rol de IAM, o
- Elimina la configuración de Bedrock en la configuración de Workload Factory.

Los diagnósticos basados en IA dejan de estar disponibles de inmediato y no se lleva a cabo ningún procesamiento de datos residual.

Perfiles de inferencia regionales e interregionales

La inferencia de Bedrock se ejecuta en la región de AWS especificada en tu perfil de inferencia. Si utilizas un perfil de inferencia entre regiones, AWS podría redirigir la solicitud a cualquier región incluida en la configuración del perfil, de acuerdo con el comportamiento estándar de AWS Bedrock. Los datos no salen de la infraestructura de AWS.

Límites de las herramientas de IA para NetApp Workload Factory

NetApp Workload Factory establece límites estrictos sobre cómo interactúan sus funciones de IA con tus entornos de AWS y ONTAP, por lo que puedes confiar en los diagnósticos asistidos por IA sin correr el riesgo de que se produzcan cambios no deseados. Amazon Bedrock utiliza herramientas de diagnóstico de AWS CLI y ONTAP CLI de solo lectura que bloquean los comandos que modifican el sistema y limitan la salida por tamaño y ejecución, y cualquier dato utilizado es solo transitorio y no es conservado ni por NetApp Workload Factory ni por AWS.

Controles de seguridad de las herramientas

Herramienta	Capacidad	Controles de seguridad
AWS CLI (solo lectura)	Ejecuta comandos de solo lectura de la CLI de AWS (describe, list, get)	Bloquea todos los verbos de modificación (crear, borrar, modificar, actualizar, añadir, eliminar). Máximo 10 comandos por paso. Salida truncada a 20 KB.
ONTAP CLI (solo lectura)	Ejecuta comandos de solo lectura de la CLI de ONTAP usando SSH	Bloquea todos los verbos de mutación (eliminar, destruir, crear, modificar, mover). Salida truncada.

El agente no puede modificar, crear ni eliminar recursos. Solo puede observar y diagnosticar.

Límites de la conservación de datos y del entrenamiento de modelos

De acuerdo con la política de AWS, Amazon Bedrock no almacena ni utiliza tus entradas y salidas para entrenar o mejorar los modelos base. En el caso de la inferencia bajo demanda (utilizada por Workload Factory), AWS procesa tus datos de forma transitoria y no los conserva tras devolver la respuesta.

Parámetros del modelo de IA y facturación para NetApp Workload Factory

NetApp Workload Factory define la configuración del modelo y los límites de uso y facturación relevantes para la gobernanza de la IA. El modelo configurado utiliza parámetros determinísticos y devuelve JSON estructurado para respaldar resultados de análisis consistentes. AWS factura la inferencia de Amazon Bedrock a tu cuenta, mientras que Workload Factory realiza un seguimiento del uso de la IA y proporciona visibilidad mensual de los costes por cuenta. Estas secciones describen los parámetros del modelo y los límites que se aplican al consumo de IA.

Modelo y parámetros

Parámetro	Valor
Modelo	Anthropic Claude (utilizando un perfil de inferencia interregional)
Temperatura	0 (determinista, sin aleatoriedad)
Número máximo de tokens de salida	2.048
Número máximo de pasos de razonamiento	15 por análisis
Formato de salida	JSON estructurado (resumen, gravedad, causa raíz, medidas correctivas)

Límites de facturación y uso

- AWS factura los servicios de inferencia de Bedrock a tu cuenta (tu perfil de inferencia, tus credenciales).
- Workload Factory realiza un seguimiento interno del uso de la IA.

- Workload Factory ofrece visibilidad de los costes mensuales por cuenta.

  "Haz un seguimiento de los costes de almacenamiento en NetApp Workload Factory"

Asistente de IA Ask Me

Arquitectura de seguridad de Ask Me para NetApp Workload Factory

Ask Me es un asistente potenciado por IA generativa integrado en NetApp Workload Factory. Ofrece asistencia conversacional en tiempo real para temas de Amazon FSx for NetApp ONTAP y Workload Factory. Las respuestas se basan en documentación verificada de NetApp, y Ask Me no puede acceder a internet pública ni usar datos de clientes para entrenar el modelo. Varias capas de seguridad ayudan a bloquear consultas maliciosas, limitar las respuestas a los dominios admitidos y evitar que la entrada del usuario sobrescriba las instrucciones del sistema. Cuando Ask Me usa herramientas, los límites de autorización, la aplicación de consultas de solo lectura, un entorno temporal de pruebas y el registro de auditoría ayudan a restringir y monitorear la ejecución.

Puedes desactivar Ask Me en cualquier momento a través de la configuración de **Administración** de Workload Factory, lo que desactiva el asistente para tu cuenta de usuario. "Obtén más información sobre cómo gestionar las funciones de GenAI."

Arquitectura RAG en Ask Me

Ask Me utiliza la generación aumentada por recuperación (RAG).

- Base de conocimientos seleccionada: las respuestas se basan en un corpus gestionado de documentación verificada y publicada de NetApp.
- Puntuación y filtrado de recuperación: solo se incluye en el contexto del modelo el contenido que resulta suficientemente relevante.
- Sin acceso a Internet: el modelo no puede recuperar, navegar ni extraer contenido externo.
- Atribución de fuentes: las respuestas incluyen referencias a los documentos fuente utilizados.

Seguridad de comandos en varias capas

- Capa 1: El clasificador de seguridad (LLM-as-a-judge) bloquea las consultas maliciosas (inyección de comandos, escalada de privilegios, exfiltración de datos, ingeniería social, incumplimientos de políticas).

  El sistema registra las consultas bloqueadas, y el modelo generativo nunca las ve.

- Nivel 2: El refuerzo de las indicaciones del sistema impide que la entrada del usuario anule las instrucciones del sistema.
- Capa 3: el ámbito del dominio limita las respuestas a los temas de FSx ONTAP y Workload Factory.
- Capa 4: El control del uso de herramientas valida los parámetros y ejecuta consultas en un entorno aislado reforzado; el modelo no puede ejecutar operaciones arbitrarias.

Aislamiento de datos de modelos y privacidad

- No se ha realizado ningún entrenamiento de modelos con datos de clientes

- Aislamiento a nivel de solicitud (sin fuga de datos entre inquilinos)
- Contexto de conversación con ámbito de sesión y historial limitado
- Sin memoria persistente de modelos
- Infraestructura de inferencia gestionada

Uso de herramientas y seguridad agénica

- Herramientas con ámbito de autorización (se aplican los límites de permisos de usuario)
- Tiempos de espera estrictos para la ejecución de herramientas
- Entorno de pruebas efímero, con ámbito de sesión para los datos recuperados
- Aplicación de consultas de solo lectura mediante allowlisting
- Auditabilidad de las llamadas a herramientas mediante el ocultamiento de parámetros confidenciales

Control de acceso de Ask Me para NetApp Workload Factory

En NetApp Workload Factory, la seguridad de acceso a Ask Me se centra en sesiones autenticadas, responsabilidad a nivel de usuario y gestión protegida de secretos durante las operaciones en tiempo de ejecución. Los controles de autenticación validan cada sesión y asocian las interacciones con un usuario específico. Las credenciales confidenciales se recuperan de un almacén de secretos gestionado en tiempo de ejecución y se ocultan en los registros. El servicio también aplica controles de infraestructura, incluyendo la ejecución de contenedores sin privilegios de root y una lista restringida de permitidos para CORS.

Autenticación

- Autenticación mediante JWT firmado con validación criptográfica (incluyendo comprobaciones de destinatario, emisor y algoritmo como RS256)
- Autenticación gestionada por middleware en el límite del servicio
- Delimitar el ámbito de identidad del usuario para que las interacciones se puedan atribuir a un usuario concreto

Protección de credenciales y gestión de secretos

- Almacenamiento seguro en un almacén de secretos: las credenciales confidenciales que utiliza el servicio se almacenan en un almacén de secretos gestionado y se recuperan en tiempo de ejecución. No se almacena ningún secreto en el código de la aplicación, en los archivos de configuración ni en las imágenes de contenedor.
- Limpieza de registros: los valores sensibles (credenciales, tokens, contraseñas, claves de acceso, certificados) se eliminan de los registros antes de que se escriban.

Seguridad de la infraestructura

- Ejecución de contenedores sin privilegios de root
- CORS restringido a una lista explícita de dominios de confianza NetApp

Modos de ejecución de Ask Me para NetApp Workload Factory

NetApp Workload Factory ofrece varios modos de ejecución para Ask Me, cada uno con diferentes características de seguridad y privacidad. Cuando Ask Me utiliza integraciones con herramientas, por ejemplo, al consultar los propios recursos de Workload Factory del cliente, la ejecución de las herramientas se controla mediante medidas de seguridad por capas. La ejecución habilitada por herramientas opera dentro de los permisos del usuario autenticado y aplica límites a la duración y al alcance de cada operación. Los datos recuperados permanecen en un entorno de pruebas de sesión efímero que bloquea el acceso externo y se destruye cuando finaliza la sesión. Los controles independientes de solicitudes y código imponen consultas de solo lectura, mientras que los registros de auditoría capturan la actividad de la herramienta con los valores sensibles ocultos.

Salvaguardas para la ejecución de herramientas

- Las herramientas con ámbito de autorización funcionan dentro de los permisos del usuario autenticado.
- El tiempo de espera de ejecución de la herramienta limita las operaciones de larga duración.
- El entorno de pruebas de datos efímeros almacena los datos recuperados por las herramientas en el ámbito de la sesión, bloquea el acceso externo, las operaciones de E/S de archivos, las llamadas de red y la carga de extensiones a nivel del motor, y se elimina cuando finaliza la sesión.
- La aplicación de consultas de solo lectura valida las consultas generadas mediante una lista de permitidos muy estricta.
- La seguridad basada en mensajes de aviso y la basada en código se aplican de forma independiente.
- La herramienta de auditabilidad registra el nombre de la herramienta, los parámetros, las marcas de tiempo y el estado del resultado, con el ocultamiento de los valores confidenciales.

Controles de salida

- Aplicación del límite de tokens
- Salida determinista para operaciones de clasificación o críticas para la seguridad (temperatura 0)
- Transmisión con finalización anticipada y limpieza

Consúltame sobre la privacidad y la disponibilidad de NetApp Workload Factory

En NetApp Workload Factory, los controles de privacidad de Ask Me limitan la exposición de los datos, mantienen aislado el procesamiento de las sesiones y preservan los límites de privacidad en las interacciones de los usuarios. Tus entradas son procesadas por un servicio de IA gestionado que no las utiliza para entrenar ni mejorar modelos base. Los datos de operaciones recuperados permanecen en un almacenamiento efímero en memoria durante toda la sesión y no se conservan ni se comparten. Una cadena de reserva multimodelo que abarca varias regiones de la nube garantiza la disponibilidad cuando el modelo principal está limitado o no está disponible, mientras que se aplican los mismos controles de seguridad en todos los modelos de reserva.

Tratamiento de datos y privacidad

- Datos de usuario en las indicaciones: procesados por un servicio de IA gestionado que no utiliza tus entradas para entrenar ni mejorar los modelos base.
- Contenido de la base de conocimientos: solo documentación seleccionada y publicada de NetApp; tus datos no están incluidos.
- Datos operativos: se recuperan solo cuando un usuario consulta sus propios recursos; se almacenan de forma efímera en la memoria durante la sesión y no se guardan ni se comparten.
- Registro de auditoría: se registran los metadatos de las consultas (ID de usuario anonimizado, marcas de tiempo, duraciones), con los campos sensibles eliminados.
- Datos de comentarios: se almacenan por separado y se vinculan a un identificador de consulta, no a información que permita identificar personalmente más allá del identificador de usuario anonimizado.

Controles de disponibilidad y aislamiento

Ask Me utiliza una cadena de respaldo multimodelo que abarca varias regiones en la nube.

Si el modelo principal se ve limitado o no está disponible, el sistema puede recurrir a modelos alternativos.

Todos los modelos están sujetos a los mismos controles de seguridad, refuerzo de las indicaciones del sistema y garantías de aislamiento.