Get started

GenAI

NetApp
October 06, 2025

This PDF was generated from https://docs.netapp.com/us-en/workload-genai/connector/quick-start-connector.html on October 06, 2025. Always check docs.netapp.com for the latest.

Table of Contents

Get started

Quick start for GenAI connectors

Get started creating a NetApp Connector for Amazon Q Business using your organization's data that exists on Amazon FSx for NetApp ONTAP file systems. After you create a connector, end users can access the Amazon Q Business assistant for organization-focused responses to their questions.

1 Log in to Workload Factory

You'll need to set up an account with Workload Factory and log in using one of the console experiences.

2 Set up your environment to meet GenAI requirements

You'll need AWS credentials to deploy the AWS infrastructure, a deployed and discovered FSx for ONTAP file system, the list of data sources that you want to integrate in your connector, access to the Amazon Q Business application, and more.

Learn more about GenAI requirements.

3 Identify the FSx for ONTAP file system that contains the data sources

The data sources that you'll integrate in your connector can be located on a single FSx for ONTAP file system, or on multiple FSx for ONTAP file systems. If these systems are in different VPCs, they must either be accessible within the same network, or the VPCs must be peered and using the same region and AWS account as the AI engine.

Learn how to identify data sources.

4 Deploy the GenAI infrastructure

Launch the infrastructure deployment wizard to deploy the GenAI infrastructure in your AWS environment. This process deploys an EC2 instance for the NetApp GenAI engine, and a volume on an FSx for ONTAP file system to contain the NetApp AI Engine databases. The volume is used to store information about the connector.

Learn how to deploy the GenAI infrastructure.

What's next

You can now create a connector for Amazon Q Business to provide organization-focused responses to end users.

GenAI connector requirements

Ensure that Workload Factory and AWS are set up properly before you create a NetApp Connector for Amazon Q Business.

Basic GenAI requirements

GenAI has general requirements that your environment needs to meet before you get started.

Workload Factory login and account

You'll need to set up an account with Workload Factory and log in using one of the console experiences.

AWS credentials and permissions

You need to add AWS credentials to Workload Factory with read/write permissions, which means you'll be using Workload Factory in *read/write* mode for GenAI.

Basic mode and *read-only* mode permissions are not supported at this time.

When setting up your credentials, selecting permissions as shown below provides you with full access to manage FSx for ONTAP file systems and to deploy and manage the GenAI EC2 instance and other AWS resources needed for your knowledge base and chatbot.

Learn how to add AWS credentials to Workload Factory

Requirements for NetApp Connector for Amazon Q Business

Ensure your environment meets the following specific requirements for Amazon Q Business connectors.

Amazon Q Business application

You need to create an Amazon Q Business application or use an existing one.

- Ensure that the application exists in one of your AWS regions.
- Ensure that you have created an index for the application.
- Ensure that the application is not in a failed state.

FSx for ONTAP file system

You need a minimum of one FSx for ONTAP file system:

- One file system will be used (or created, if it doesn't exist) by the NetApp GenAI engine to store information about the connector.

  This FSx for ONTAP file system must use FlexVol volumes. FlexGroup volumes are not supported.

- One or more file systems will contain the data sources that you'll be adding to your connector.

  One FSx for ONTAP file system can be used for both of these purposes, or you can use multiple FSx for ONTAP file systems.

- You'll need to know the AWS region, VPC, and subnet where the AWS FSx for ONTAP file system resides.
- You'll need to consider the tag key/value pairs that you want to apply to the AWS resources that are part of this deployment (optional).
- You'll need to know the key pair information that allows you to securely connect to the NetApp AI engine instance.

Learn how to deploy and manage FSx for ONTAP file systems

Identify data sources to add to a connector

Identify, or create, the documents (data sources) that reside on your FSx for ONTAP file system that you'll integrate in your connector. These data sources enable Amazon Q Business to provide accurate and personalized answers to user queries based on data that is relevant to your organization.

Maximum number of data sources

The maximum number of supported data sources is 10.

Location of data sources

Data sources can be stored in a single volume, or in a folder within a volume, on an SMB share or NFS export on an Amazon FSx for NetApp ONTAP file system. Data sources can also be stored on Amazon FSx for NetApp ONTAP volumes that are in a NetApp SnapMirror data protection relationship.

You can't select individual documents within a volume or folder, therefore, you should ensure that each volume or folder that contains data sources does not contain extraneous documents that shouldn't be integrated with your knowledge base.

You can add multiple data sources into each connector, but they all need to reside on FSx for ONTAP file systems that are accessible from your AWS account.

The maximum file size for each data source is 50 MB.

Supported protocols

Connectors support data from volumes that use either NFS or SMB/CIFS protocols. When selecting files stored using the SMB protocol, you'll need to enter the Active Directory information so that the connector can access the files on those volumes. This includes the Active Directory Domain, IP address, user name, and password.

When storing your data source on a share (file or directory) accessed over SMB, the data is only accessible by chatbot users or groups who have the permissions to access that share. When this "permission-aware capability" is enabled, the AI system will compare the user email in auth0 to the users allowed to view or use the files on the SMB share. The chatbot will provide answers based on user permissions for the embedded files.

For example, if you have integrated 10 files (data sources) into your connector, and 2 of the files are human resources files that contain restricted information, only chatbot users who are authenticated to access those 2 files will received responses from the chatbot that include data from those files.

When you add data sources to an Amazon Q Business connector, only user permissions apply to data source files. Group permissions are not applied.

If a file in your data source lacks text (for example, a text-free image), Amazon Q Business does not index it but logs an entry in Amazon CloudWatch Logs noting the absence of text.

Supported data source file formats

The following data source file formats are currently supported with NetApp Connector for Amazon Q Business.

File format	Extension
Comma-separated values file	.csv
JSON and JSONP	.json
Markdown	.md
Microsoft Word	.docx
Plain text	.txt
Portable Document Format	.pdf
Microsoft PowerPoint	.ppt or .pptx
Hypertext Markup Language	.html
Extensible Markup Language	.xml
XSLT	.xslt
Microsoft Excel	.xls
Rich Text Format	.rtf

Deploy the GenAI infrastructure

Unresolved directive in connector/deploy-infrastructure.adoc - include::_include/deploy-infrastructure.adoc[]

Copyright information

Copyright © 2025 NetApp, Inc. All Rights Reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP “AS IS” AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

LIMITED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (b)(3) of the Rights in Technical Data -Noncommercial Items at DFARS 252.227-7013 (FEB 2014) and FAR 52.227-19 (DEC 2007).

Data contained herein pertains to a commercial product and/or commercial service (as defined in FAR 2.101) and is proprietary to NetApp, Inc. All NetApp technical data and computer software provided under this Agreement is commercial in nature and developed solely at private expense. The U.S. Government has a non-exclusive, non-transferrable, nonsublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b) (FEB 2014).

Trademark information

NETAPP, the NETAPP logo, and the marks listed at http://www.netapp.com/TM are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.