200 XP ▶

How the Azure Data Box family works

7 minutes

After looking at the options for moving data, you decide that Azure Data Box Disk is the best choice. You don't want to dedicate your entire network bandwidth to the task of importing your data to Azure. As the administrator for your organization, you need to better understand the overall workflow involved in getting the data into Azure.

In this unit, you'll learn how to use Azure Data Box Disk and the other devices in the Azure Data Box family to import data to Azure.

Import data by using Azure Data Box family

The following diagram shows the high-level steps for ordering a device from the Azure Data Box family:
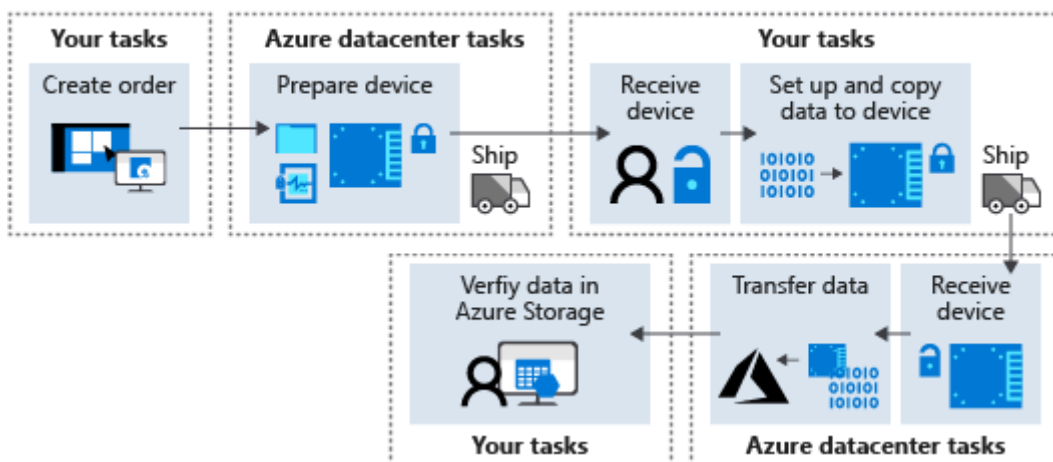
For clarity, let's break out these workflow stages into discrete steps:

1. Create an order in the Azure portal or by using Azure CLI or PowerShell.
2. Receive the device from the Azure datacenter. Connect to and unlock the device.
3. Set up and copy data to the device. Validate the data.
4. Ship the device to Azure.
5. Verify the data in Azure storage.

If you plan to order Data Box Heavy, make sure that it will fit through the entrance of your premises, and that you have at least two people to handle it. It's a massive device that weighs around 500 lbs and arrives on wheels.

We'll cover the specific steps to use Azure Data Box Disk later in this unit.

Copy data using standard tools

You can copy data by using standard tools. For example, use File Explorer to drag and drop files. Or, use any Server Message Block (SMB)-compatible file-copy tool like Robocopy.

Use Azure Storage file-naming conventions and size limits

When you copy data, all the standard Azure Storage naming conventions apply:

- Subfolder names should be lowercase, from 3 to 63 characters, and consist only of letters, numbers, and hyphens. Consecutive hyphens aren't allowed.
- Directory and file names for Azure Files shouldn't exceed 255 characters.
- File size must not exceed ~4.75 tebibytes (TiB) for block blobs, ~8 TiB for page blobs, and ~1 TiB for Azure Files.

Copy data into the appropriate folder for your storage type: PageBlob, BlockBlob, AzureFile, or ManagedDisk.

- Use the ManagedDisk folder for virtual hard disks (VHDs) that you want to migrate to Azure. Use the PageBlob folder for VHDX files.
- Any files copied directly to the PageBlob or BlockBlob folders are inserted in a default **$root** container. Subfolders are created as containers in Azure.
- For Azure Files, files must be in subfolders under the AzureFile folder. Any files copied to the root of the AzureFile folder are uploaded as block blobs instead of Azure Files items.

If you don't follow the file structure, size limit, and naming conventions, the data upload to Azure might fail. If you're using Windows, we recommend that you validate the files by using DataBoxDiskValidation.cmd, which is provided in the DataBoxDiskImport folder. If you have time, use the *generate checksums* option to validate your data before sending it to Azure.

Import data by using Azure Data Box Disk

For your vehicle data, you'd follow these steps to order a Data Box Disk to import the data to Azure:

1. **Create an order**

   - Start by making an assessment of how much data your vehicles generate during the time you plan to sample before shipping it to Azure.
   - Create an order in the portal or by using Azure CLI or PowerShell. Specify the amount of data, source country, and destination region.

2. **Receive the device from Azure Datacenter**

- The service assigns from one to five 8-TB BitLocker-encrypted SSDs (up to 35 TB of usable capacity) and ships them within 10 days. If you need more than 35 TB, you can create more orders.
- When you receive the disks, sign in to the Azure portal and go to the **Device Details** page for Data Box Disk. Get the passkey. Download the right version of the Data Box Disk unlocker tool for your operating system.

3. **Set up and copy data to the device**

- For the autonomous vehicle scenario, you connect the disks directly to the onboard Linux or Windows computer in the car. With Azure Data Box Disk, you can connect multiple disks simultaneously and run multiple copy jobs in parallel.
- Azure Data Box Disk supports Azure block blob, Azure page blob, Azure Files, and Azure Managed Disks storage types for upload to Azure. The disks come with a predefined folder structure, based on the storage type. Copy data into the appropriate folder for your storage type: PageBlob, BlockBlob, AzureFile, or ManagedDisk.
- Validate the files by running DataBoxDiskValidation.cmd, which is provided in the DataBoxDiskImport folder. Select option 1 to validate the files. If you have time, we recommend that you generate checksums by selecting option 2. Generating a checksum may take some time, depending on the data size.
- The first three days are billed as a single day, as a grace period, plus one-off shipping and order processing fees. After the first three days, there's a per-day cost for each SSD.

4. **Ship device to Azure datacenter**

Ship the disks back to Azure, using the provided shipping label for your regional carrier. When the device is received by the Azure datacenter, the data is copied to your Azure storage account.

5. **View data in Azure Storage**

Check for any errors, and verify that the data has been copied to your storage account.

Clone a Data Box order

Because you'll need to periodically ship telemetry data from the autonomous vehicles, you'll need an easy way to repeat the process described here. The best way is to clone the previous Azure Data Box Disk order in the portal. When you select **Clone** within an order, all the details of the previous order remain the same, and the name becomes the original order name but appended with **-Clone**.

Export data by using Data Box

You can export data from Azure by using Data Box. Data Box Disk and Data Box Heavy don't support exporting data from Azure.

Though we won't need to export data from Azure for our scenario, you should be aware of the workflow. The steps to order and use Data Box to export data is similar to the import workflow. At a high level, the workflow typically includes the following steps:

1. Create an order in the Azure portal.
2. Receive the device from the Azure datacenter. Connect to and unlock the device. The device will have data that the Azure datacenter copied from your Azure storage account.
3. Copy data from the device.
4. Return the device to Azure. The Azure datacenter securely erases the device disks as per the National Institute of Standards and Technology (NIST) guidelines.

Check your knowledge

1. What method would you use to unlock the disks you've received?

○ Get the passkey from the **Device Details** page in the Azure portal.

○ Look on the reverse side of the Data Box Disk device for the passkey.

○ Use the passkey in the root folder of the Data Box Disk device.

2. You want to copy a virtual hard disk in the VHDX format to an Azure Data Box Disk. What folder do you copy the VHDX to?

○ PageBlob folder

○ BlockBlob folder

○ AzureFiles folder

○ ManagedDisk folder

Check your answers