

200 XP



# When to use Azure Data Box family

7 minutes

Azure Data Box family devices aren't the only ways to get data into Azure. For your autonomous vehicle data, make sure that Data Box Disk is your best choice before you place an order.

In this unit, you'll learn when to use a Data Box device and when to use some other solution.

## Use cases for Data Box devices

Data Box is ideally suited to transfer terabytes of data in scenarios with no to limited network connectivity. The data movement can be one-time, periodic, or an initial bulk data transfer followed by periodic transfers.

Here are the various scenarios where Data Box devices can be used to import data to Azure.

- **One time migration** - when large amount of on-premises data is moved to Azure.
  - Moving a media library from offline tapes into Azure to create an online media library.
  - Migrating your VM farm, SQL server, and applications to Azure
  - Moving historical data to Azure for in-depth analysis and reporting using HDInsight
- **Initial bulk transfer** - when an initial bulk transfer is done using Data Box (seed) followed by incremental transfers over the network.
- **Periodic uploads** - when large amount of data is generated periodically and needs to be moved to Azure. For example in energy exploration, where video content is generated on oil rigs and windmill farms.

Here are the various scenarios where Data Box can be used to export data from Azure. Data Box Disk and Data Box Heavy don't support exporting data from Azure.

- **Disaster recovery** - when a copy of the data from Azure is restored to an on-premises network. In a typical disaster recovery scenario, a large amount of Azure data is exported to a Data Box. Microsoft then ships this Data Box, and the data is restored on your premises in a short time.
- **Security requirements** - when you need to be able to export data out of Azure due to government or security requirements. For example, Azure Storage is available in US Secret and Top Secret clouds, and you can use Data Box to export data out of Azure.

- **Migrate back to on-premises or to another cloud service provider** - when you want to move all the data back to on-premises, or to another cloud service provider, export data via Data Box to migrate the workloads.

## Offline transfer options

The offline options to move your data to Azure include Data Box devices and Azure Import/Export.

### Data Box devices

To get data into Azure, you can load your data onto physical devices and send them to an Azure datacenter. There, the data is copied into your Azure subscription. As you saw in unit 2, three Data Box products work this way:

- Data Box
- Data Box Disk
- Data Box Heavy

If you're not in a region supported by Azure Data Box family, consider using Azure Import/Export to import data into Azure.

### Azure Import/Export

The other service that uses physical devices to transfer data into Azure is Azure Import/Export. Azure Import/Export allows you to use your own disks. You can do offline transfers of data from Blob storage to disks. Azure Import/Export also exports data from Azure back to your own datacenters and allows data to cross international borders.

The setup procedure to import data using Azure Import/Export is more difficult than Azure Data Box family. You have to handle all the shipping details and device preparation. We recommend using Azure Data Box, a done-for-you service that minimizes the effort required to move huge volumes of data inexpensively and securely into Azure.

## Network transfer options

Another way to get data into Azure is to send it across the internet. There are several tools you can use to transfer data across the network.

### Data Box Gateway

As discussed in unit 2, Data Box Gateway is a virtual appliance. It's based on a virtual machine that you provision in your on-premises environment. You write data to the virtual device by using the NFS and SMB protocols. The device then transfers your data to Azure block blob, page blob, or Azure Files.

## Azure Storage Explorer

Azure Storage Explorer is a free, downloadable application that you can run on Windows, Mac, or Linux. It provides a graphical user interface (GUI) that you can use to explore blobs, files, queues, and tables in Azure storage accounts. If you transfer only a few files at a time and you don't need to automate transfers, Storage Explorer is an ideal solution.

If you don't want to install an application, but you have a browser, you can use the Storage Explorer page in the Azure portal to transfer files.

## Azure Stack Edge

Azure Stack Edge also uses physical devices to transfer data into Azure. Unlike Data Box devices, Azure Stack Edge devices provide Azure capabilities like compute, storage, networking, and hardware-accelerated machine learning to any edge location. Use an Azure Stack Edge device for the following scenarios:

- **Inference with Azure Machine Learning** - Run ML models to get quick results that can be acted on before the data is sent to the cloud. The full data set can optionally be transferred to continue to retrain and improve your ML models.
- **Preprocess data** - Transform data via compute options such as containers or virtual machines before sending it to Azure to create a more actionable dataset. Preprocessing can be used to:
  - Aggregate data.
  - Modify data, for example to remove personal data.
  - Subset data to optimize storage and bandwidth, or for further analysis.
  - Analyze and react to IoT Events.
- **Transfer data over network to Azure** - Easily and quickly transfer data to Azure to do further compute and analytics or for archival purposes.

## Azure Data Factory

Azure Data Factory is a service that enables you to organize, move, and transform large quantities of data from many different sources. In Data Factory, you create data pipelines that ingest data from relational databases, NoSQL databases, and other systems. You can use Azure

Machine Learning, Hadoop, Spark, and other services to process and transform that data. Then, at the end of the pipeline, you can publish the transformed data to Azure SQL Data Warehouse, Azure SQL Database, Azure Cosmos DB, and Azure Storage.

Use this service, if you have complex data transformation needs but don't want to write scripts or compile code.

## Scripted or programmatic transfer

Microsoft provides a range of optimized command-line tools, which you can call with individual commands or assemble into scripts, and which can move data into Azure. These tools include:

- **AzCopy.** You can call this command-line program from any script to transfer blobs or files to an Azure storage account.
- **Azure PowerShell.** PowerShell is a scripting interface and language designed to help administrators automate complex tasks by running commands called **cmdlets**. The Azure PowerShell module provides a range of cmdlets to work with Azure. These include cmdlets that you can use to upload data into Azure storage accounts, Azure SQL Database, Azure Cosmos DB, and other destinations. For example, to upload a file to an Azure Storage blob, use the `Set-AzStorageBlobContent` cmdlet.
- **Azure CLI.** The Azure CLI is a command-line tool that you can call from PowerShell and Bash scripts to work with Azure. It includes options that can move data into Azure locations. For example, use the command `az storage blob upload` to upload a file to an Azure Storage blob.

Scripting tools are easy for administrators and non-developers to use. If you have people with developer skills who are able to write compiled code, you can consider using programming languages to orchestrate data transfer into Azure. Your code must call the Azure Storage Representational State Transfer (REST) application programming interfaces (APIs). The REST APIs are a set of web services that you can call from many different programming languages. Developers can use .NET languages, Java, Python, Node.js, C++, PHP, among other languages, to call these APIs. They'll choose the one they're most familiar with.

## Choosing a data transfer method

To choose a transfer method from these systems, consider the following questions:

- What's the size of the data?
- How often do I want to transfer data?
- How much network bandwidth is available?

If you want to transfer a few files occasionally, choose Azure Storage Explorer or use the Azure portal.

If network bandwidth is low or expensive to use, and you have a large volume of data, use a physical device to transfer data. If you want to use your own disks, choose Azure Import/Export. If you prefer the convenience of using Microsoft hardware, choose an Azure Data Box product. Choose from Data Box, Data Box Disk, and Data Box Heavy according to their maximum data volumes.

If you have plenty of available network bandwidth and you want to transfer data regularly or continuously, choose a network transfer method. If you want to transform the data as you transfer it, without using code, consider Azure Data Factory. If you can invest the time and have the skills to write code, use scripting tools or developed compiled code against the REST APIs. If you want to transfer large datasets over the network and you have high available bandwidth, choose Data Box Gateway or an Azure Stack Edge device.

## Data transfer for the autonomous vehicles

In your organization, you want to upload autonomous vehicle telemetry data to Azure in a fast, relatively cheap, and secure way without hogging the network. That's a key use case for Azure Data Box Disk.

You can use Azure Data Box Disk for different kinds of data transfer scenarios. Azure Data Box Disk is suitable for one-time migrations like moving up to 35 TB of data from offline tapes to Azure cool storage. If there's more than 35 TB to upload, you can create extra orders. It's also worth reviewing the other Data Box family options that you explored in the first unit, as there might be a more suitable candidate. Incremental transfers also work well. Data Box Disk is used to provide the initial, large seed, after which further increments of data are copied by using standard network techniques.

## Check your knowledge

1. Which data import method is best for importing daily traffic-camera video data when you have moderate to high network bandwidth?

- ☐ Data Box Gateway
- ☐ Azure Stack Edge
- ☐ Data Box Disk

2. What's the maximum amount of data that can be transferred to Azure in one operation through the Azure Data Box Disk?

- ☐ 35 TB
- ☐ 500 TB
- ☐ 80 TB

3. You have a huge amount of data being generated by smart devices and applications at your data center. You want to perform rapid machine learning-based inference on the data before moving it to Azure for deeper analysis. Which device would you use?

- ☐ Azure Stack Edge device
- ☐ Data Box Gateway
- ☐ Data Box Heavy

Check your answers

---