

## Lab 1: Loading series of file according to LastModified

This lab guides you through loading files from Azure File Storage to ADLSgen2, based on the source files' LastModified attribute.


### Step 1: Create an ADLSgen2 as Destination Folder

- On Azure portal, [New Service], then search for Storage Account, [Create]

---

**Storage account**  
Microsoft

---



**Storage account**  [Save for later](#)  
Microsoft  
[Create](#)

---

Microsoft Azure provides scalable, durable cloud storage, backup, and recovery solutions that effectively enhance your existing applications and business continuity strategies. Store binary data such as video, audio, and images.

Useful Links

- [Documentation](#)
- [Service overview](#)
- [Pricing](#)

- Define below attributes for the ADLSgen2, do note the Storage Account Name has to be unique

Basics Advanced Tags Review + create

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more](#)

### Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

* Subscription	Visual Studio Enterprise
* Resource group	(New) storageRG

[Create new](#)

### Instance details

The default deployment model is Resource Manager, which supports the latest Azure features. You may choose to deploy using the classic deployment model instead. [Choose classic deployment model](#)

* Storage account name ⓘ	gen2demo321
* Location	(Asia Pacific) East Asia
Performance ⓘ	<input checked="" type="radio"/> Standard <input type="radio"/> Premium
Account kind ⓘ	StorageV2 (general purpose v2)
Replication ⓘ	Locally-redundant storage (LRS)
Access tier (default) ⓘ	<input checked="" type="radio"/> Cool <input type="radio"/> Hot

- And in the Advanced Tab:

[Basics](#)
[Advanced](#)
[Tags](#)
[Review + create](#)

---

**Security**

Secure transfer required ⓘ ☐ Disabled ☒ Enabled

**Virtual networks**

Allow access from ☒ All networks ☐ Selected network  
 ⓘ All networks will be able to access this storage account. [Learn more](#)

**Data protection**


Blob soft delete ⓘ ☐ Disabled ☐ Enabled  
 ⓘ Blob soft delete and hierarchical namespace cannot be enabled simultaneously.

**Data Lake Storage Gen2**

Hierarchical namespace ⓘ ☐ Disabled ☒ Enabled

- And click though [Next>] in Tags, then [Create]
- Once the resource creation is done, click on the resource name (or enter the resource name in the search bar). Click on below button to enter ADLSgen2 configuration page:

## Services



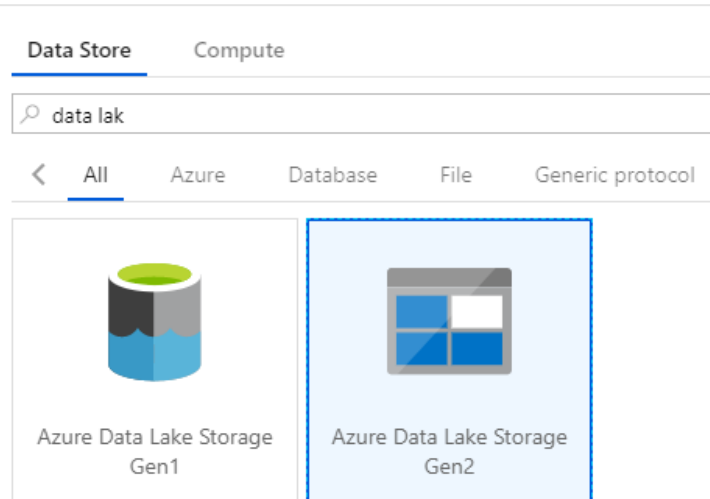
**Data Lake Gen2 file systems**  
 Massively scalable data lake storage  
[Learn more](#)

- [+ File System] to create a new file system. Enter [adlsfs] to create a new file system. Your ADLSgen2 is ready to use.

## Step 2: Setup the connection on ADF for ADLSgen2

- In [Connections], Search for ADLSgen2 and click Continue

### New Linked Service

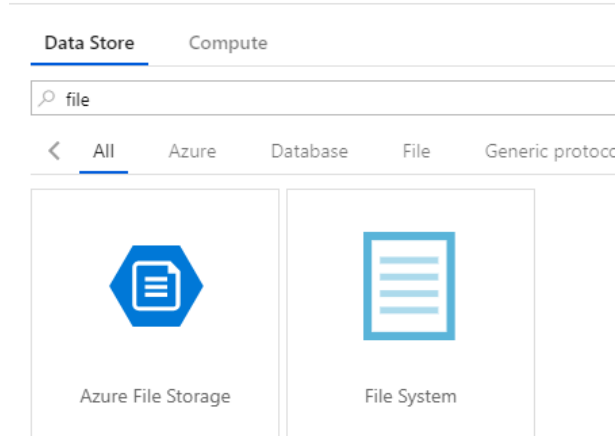


- Click through to search for your own ADLSgen2 by account key. If you're using the same subscription, it shows up automatically.

### Step 3 Create Azure File Storage connection

- Click [Connections], Then New

#### New Linked Service

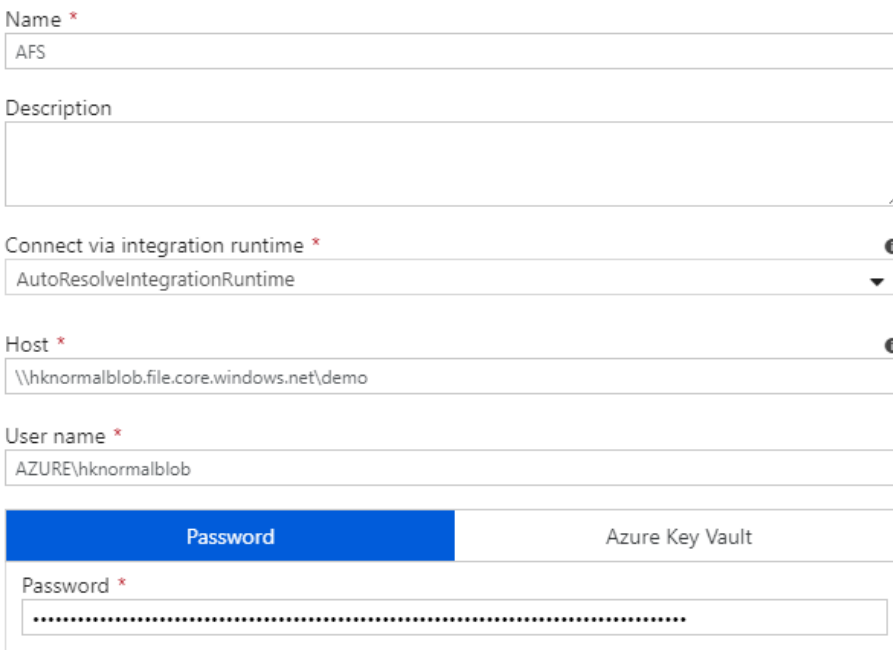


- Fill in per below details
  - In the Password, fill in the Access key

iMSyy7440wasS0Jk74Irb8iYptfv8yaVgb/OjS2Y6nlh4oL16bH1yQ0umLmHeBwG7ioniMa+EK+MsOqkNZDJ/  
w==

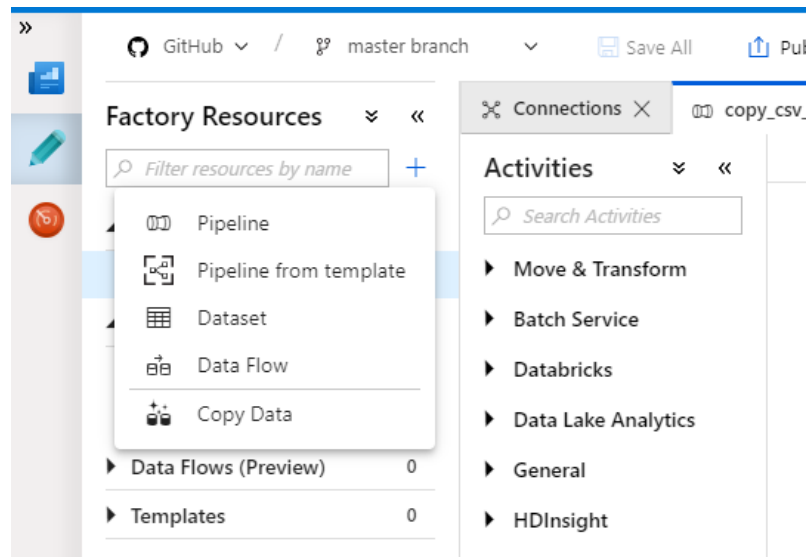
- Host : [\\hknormalblob.file.core.windows.net\demo](https://hknormalblob.file.core.windows.net/demo)
- Username : AZURE\hknormalblob

#### ← New Linked Service (Azure File Storage) ×



### Step 3 Create a Copy Data activity pipeline

- Click [+] button under Factory Resource, select "Copy Data"



- Select New Dataset, and type in file for file storage
- For [Source Data store],

## Choose the input file or folder

Select a source file or folder to be copied to the destination data store.

File or folder	<input type="text"/>	
Binary copy	<input checked="" type="checkbox"/>	<a href="#">?</a>
Compression type	<input type="text" value="None"/>	
Recursively	<input checked="" type="checkbox"/>	<a href="#">?</a>
Max concurrent connections	<input type="text"/>	<a href="#">?</a>
Filter by last modified	Start time (UTC) <input type="text"/>	End time (UTC) <input type="text"/>

- Select Destination data store

## Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

All
Azure
Database
File
Generic protocol
NoSQL
Services and apps

All
Filter by name
+ Create new connection

AFS

gen2storagedemo

MikeADLSgen2

- Specify a meaningful Folder path like adlsngen2/demodest
- Keep the settings as default

## Settings

More options for data movement

### Performance settings

Enable staging ☐ ⓘ

### Advanced Settings

Data integration unit 

Auto ⓘ

  
☐ Edit

You will be charged # **of used DIUs** \* **copy duration** \* **\$0.25/DIU-hour** (local currency and separate discounting may apply per subscription type). When you choose "Auto", Data Factory dynamically applies the optimal DIU setting based on your source-sink pair and data pattern. See [here](#) for current prices and learn how to optimize copy by DIU setting from [here](#).

Degree of copy parallelism 

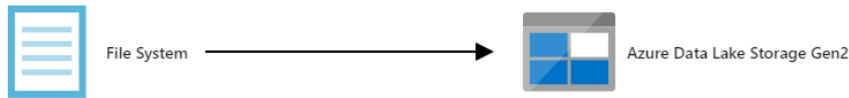
Auto ⓘ

  
☐ Edit

- You will have similar Summary

## Summary

You are running pipeline to copy data from File System to Azure Data Lake Storage Gen2.



### Properties

[Edit](#)

Task name Copy\_AFS\_2\_ADLS

Task description

### Source

[Edit](#)

Connection name AFS

Dataset name SourceDataset\_28u

File name

Folder path

### Destination

[Edit](#)

Connection name MikeADLSgen2

Dataset name DestinationDataset\_28u

File name

Folder path adlsgen2/demodest

### Copy settings

[Edit](#)

[Previous](#)

[Next](#)

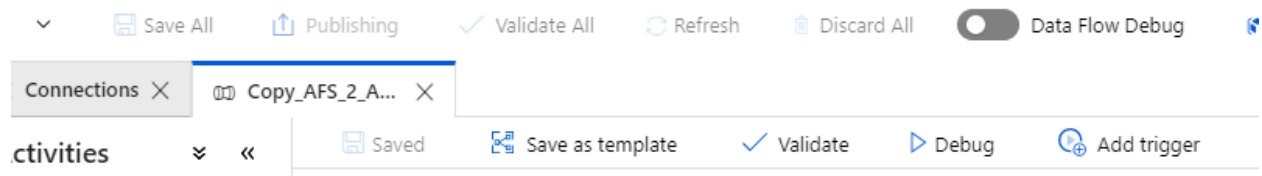
## Step 4 Execute a non-filtered copy

Locate your saved pipeline, [Copy\_AFS\_2\_ADLS] in ADF design portal

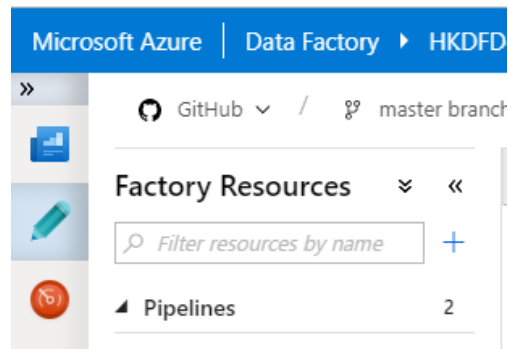
The screenshot shows the ADF design portal interface. On the left, the 'Factory Resources' pane lists 'Pipelines' and 'Datasets'. The 'Copy\_AFS\_2\_ADLS' pipeline is selected. The main area shows the 'Activities' pane with a search bar and a list of activity categories: Move & Transform, Batch Service, Databricks, Data Lake Analytics, and General. The 'Copy Data' activity is visible in the canvas, with the dataset 'Copy\_28u' selected.



- Click [Publish] to commit the change



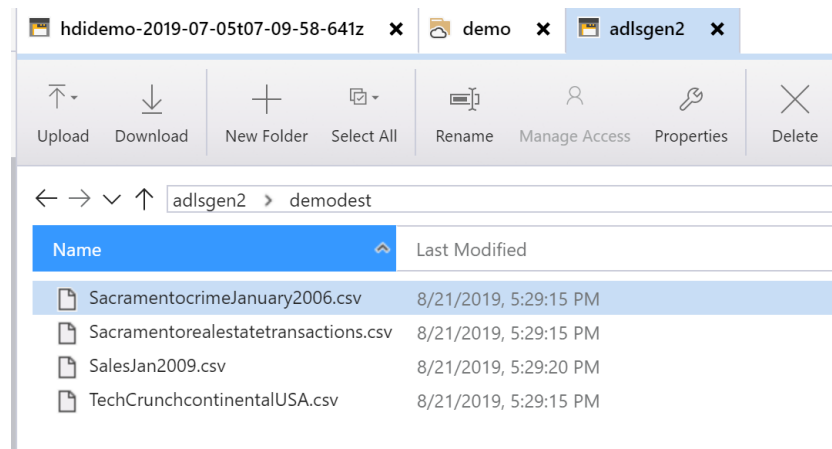
- Then go for [Add trigger] => [Trigger Now], you're going to execute the pipeline on demand. We go to the Monitor page to monitor the status



- Once it's finished, you will find the STATUS set as Succeeded,

Last 24 Hours 08/20/2019 5:29 PM - 08/21/2019 5:29 PM		Time Zone (UTC+08:00) Beijing, Chongqing, Hon...		View All Rerun History		Gantt View	
All	Succeeded	In Progress	Queued	Failed	Cancelled		
PIPELINE NAME	ACTIONS	RUN START	DURATION	TRIGGERED BY	STATUS	PARAMETERS	ANNOTATIONS
Copy_AFS_2_ADLS		08/21/2019, 5:28:56 PM	00:00:27	Manual trigger	<span style="color: green;">✔ Succeeded</span>		

- And files are visible to user in Storage Explorer.




### Step 5 Execute a filtered copy

- Now, delete all **destination** folder files. Click on the Copy Data activity, in the Modified datetime start, define as 08/21/2019 08:00:00

Modified datetime start (UTC)

- Remember, every change needs to be published.
- Once executed, you will find that only 1 file is copied.

← → ∨ ↑	adlsngen2 > demodest
Name	Last Modified
 SalesJan2009.csv	8/21/2019, 5:51:06 PM

### Step 6 dynamic filter criteria

In the real world, you won't want a hardcoded value. Now if you go back to Copy Data Activity, you can find the [Add dynamic content]





Wildcard folder path

Wildcard file name

Modified datetime start (UTC)

[Add dynamic content \[Alt+P\]](#)

- Now, you can implement a regular expression based on system variables

General	Source	Sink	Mapping	Settings	User Properties
Source dataset *	 SourceDataset_28u	 Edit	 New		
Recursively	<input checked="" type="checkbox"/> 				
Wildcard folder path	<input type="text" value=""/>				
Wildcard file name	<input type="text" value="*.csv"/>				
Modified datetime start (UTC)	<input type="text" value="@{addhours(utcnow(),-3)}"/>				