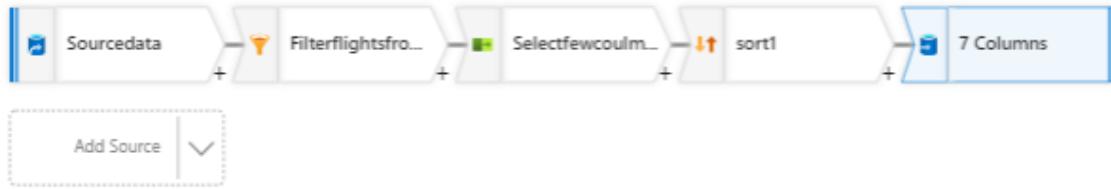


## Azure Data Factory Project:

**Dataflow project** – Created a Dataflow to Filter, Select and sort by year and save in the storage



1. Taken a dataset of US flights from kaggle see screenshot below.

US Airline Flight Routes and Fares 1992-2024 - Excel (Product Activation Failed)																										
File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do										Conditional Formatting Table Styles										Insert Delete Format						
Calibri 11 A A										General Normal Bad Good Neutral Input										Autofill Clear						
Paste Copy Format Painter										Calculation check cell Explantory																
Clipboard Font Alignment Number																				Cells						
D46 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z																										
tbl	Year	quarter	citymarketid_1	citymarketid_2	city1	city2	airportid	airportname	airport_2name	passengers	carrier	large_msfare	carrier_1	carrier_2	fare_low	Geocode	Geocode_2	tblback								
2	Tabelle	2021	3	30135	33195	Allentown/Bethlehem/Etapa, FL (Metro)	10135	14112	ABE	PIE	970	180	81.43	G4	1	81.43			2021031951412ABEPIE							
3	Tabelle	2021	3	30135	33195	Allentown/Bethlehem/Etapa, FL (Metro)	10135	15304	ABE	TPA	970	19	208.99	DL	0.4659	219.98	UA	0.1193	254.11	2021031951412ABETPA						
4	Tabelle	2021	3	30140	30194	Albuquerque, NM	10140	11239	ABQ	DAL	580	204	184.56	WN	0.9968	184.44	WN	0.9968	184.44	20210314011239ABQDAL						
5	Tabelle	2021	3	30140	30254	Albuquerque, NM	10140	11298	ABQ	DFW	580	264	182.64	AA	0.9774	183.09	AA	0.9774	183.09	20210314011298ABQDFW						
6	Tabelle	2021	3	30140	30486	Albuquerque, NM	10140	14107	ABQ	PHX	328	398	177.11	WN	0.6061	184.49	AA	0.9399	165.77	20210314011407ABQPHX						
7	Tabelle	2021	3	30140	30721	Albuquerque, NM	10140	10721	ABQ	BOS	1974	153	324.97	AA	0.4283	323.73	WN	0.1609	298.2	202103140110721ABQBOS						
8	Tabelle	2021	3	30140	30721	Albuquerque, NM	10140	11296	ABQ	MHT	1974	16	315.9	WN	0.7285	270.42	WN	0.7285	270.42	202103140111296ABQMHT						
9	Tabelle	2021	3	30140	30721	Albuquerque, NM	10140	14307	ABQ	PVD	1974	22	329.22	WN	0.5415	271.6	WN	0.5415	271.6	202103140114307ABQPVD						
10	Tabelle	2021	3	30140	30852	Albuquerque, NM	10140	10821	ABQ	BWI	1670	159	255.89	WN	0.7212	244.89	WN	0.7212	244.89	202103140110821ABQBWI						
11	Tabelle	2021	3	30140	30852	Albuquerque, NM	10140	11278	ABQ	DCA	1670	151	291.18	AA	0.4404	296.88	WN	0.3197	247.2	202103140111278ABQDCA						
12	Tabelle	2021	3	30140	30852	Albuquerque, NM	10140	10264	ABQ	IAD	1670	59	343.58	UA	0.5646	382.06	WN	0.1402	266.61	202103140110264ABQIAD						
13	Tabelle	2021	3	30140	30977	Albuquerque, NM	10140	11332	ABQ	MDW	1121	99	231.66	WN	0.9934	230.85	WN	0.9934	230.85	202103140111332ABQMDW						
14	Tabelle	2021	3	30140	30977	Albuquerque, NM	10140	13930	ABQ	ORD	1121	206	241.25	AA	0.6889	229.41	AA	0.6889	229.41	202103140113930ABQORD						
15	Tabelle	2021	3	30140	31453	Albuquerque, NM	10140	11191	ABQ	HOU	759	166	203.93	WN	0.9712	203.44	WN	0.9712	203.44	202103140111191ABQHOU						
16	Tabelle	2021	3	30140	31453	Albuquerque, NM	10140	12266	ABQ	IAH	759	148	232.85	UA	0.8512	235.69	WN	0.0447	213.54	202103140112266ABQIAH						
17	Tabelle	2021	3	30140	31703	Albuquerque, NM	10140	11618	ABQ	EWR	1861	77	342.13	UA	0.3845	382.33	AA	0.3324	296.4	202103140111618ABQEWR						
18	Tabelle	2021	3	30140	31703	Albuquerque, NM	10140	12197	ABQ	HPN	1861	3	356.92	DL	0.6154	358.24	AA	0.3462	305.22	202103140112197ABQHPN						
19	Tabelle	2021	3	30140	31703	Albuquerque, NM	10140	12391	ABQ	ISP	1861	4	240.1	WN	0.8049	220.94	WN	0.8049	220.94	202103140112391ABQISP						
20	Tabelle	2021	3	30140	31703	Albuquerque, NM	10140	12478	ABQ	JKF	1861	177	251.78	BE	0.8338	234.07	BE	0.8338	234.07	202103140112478ABQJKF						
21	Tabelle	2021	3	30140	31703	Albuquerque, NM	10140	12553	ABQ	LGA	1861	125	295.22	AA	0.5013	302.37	WN	0.2145	229.04	202103140112553ABQLGA						
22	Tabelle	2021	3	30140	31703	Albuquerque, NM	10140	15070	ABQ	SWF	1861	0	427.5	AA	1	427.5	AA	1	427.5	202103140115070ABQSWF						
23	Tabelle	2021	3	30140	32457	Albuquerque, NM	10140	13796	ABQ	OAK	896	216	194.36	WN	0.9638	193.2	WN	0.9638	193.2	202103140113796ABQOAK						
24	Tabelle	2021	3	30140	32457	Albuquerque, NM	10140	14771	ABQ	SFO	896	188	246.07	UA	0.6493	255.54	WN	0.1346	221.69	202103140114771ABQSFO						
25	Tabelle	2021	3	30140	32457	Albuquerque, NM	10140	14611	ABQ	SJC	896	61	209.22	WN	0.6566	200.07	WN	0.6566	200.07	202103140114611ABQSJC						
26	Tabelle	2021	3	30140	32575	Albuquerque, NM	10140	10800	ABQ	BUR	677	39	205.45	WN	0.6592	205.43	AA	0.243	196.01	202103140110800ABQBUR						
27	Tabelle	2021	3	30140	32575	Albuquerque, NM	10140	13892	ABQ	LAX	677	572	172.99	DL	0.379	181.73	WN	0.2572	158.14	202103140113892ABQLAX						
28	Tabelle	2021	3	30140	32575	Albuquerque, NM	10140	12954	ABQ	LSB	677	12	201.28	WN	0.5636	187.75	DL	0.1182	180.77	202103140112954ABQLSB						
29	Tabelle	2021	3	30140	32575	Albuquerque, NM	10140	13891	ABQ	ONT	677	52	212.31	WN	0.5993	204.21	WN	0.5993	204.21	202103140113891ABQONT						
30	Tabelle	2021	3	30140	32575	Albuquerque, NM	10140	14908	ABQ	SNA	677	79	222.87	WN	0.614	216.43	AA	0.3201	215.59	202103140114908ABQSNA						
31	Tabelle	2021	3	30154	31703	Nantucket, MA	10154	11618	ACK	EWR	218	176	185.27	UA	0.5473	189.12	BE	0.446	180.08	202103141511618ACKEWR						
32	Tabelle	2021	3	30154	31703	Nantucket, MA	10154	11197	ACK	HPN	218	78	189.08	BE	0.9958	188.57	BE	0.9958	188.57	202103154121197ACKHPN						
33	Tabelle	2021	3	30154	31703	Nantucket, MA	10154	12478	ACK	JKF	218	210	156.66	BE	0.7464	155.54	BE	0.7464	155.54	2021031541212478ACKJKF						
34	Tabelle	2021	3	30154	31703	Nantucket, MA	10154	12553	ACK	LGA	218	135	157.09	BE	0.846	153.24	BE	0.846	153.24	2021031541212553ACKLGA						
35	Tabelle	2021	3	30189	30194	Colorado Springs, CO	11189	11259	COS	DAL	602	194	122.76	WN	1	122.76	WN	1	122.76	2021031911259COSDAL						
36	Tabelle	2021	3	30189	30486	Colorado Springs, CO	11189	11298	COS	DFW	602	343	154.47	AA	0.98	153.63	AA	0.98	153.63	2021031911298COSDFW						
37	Tabelle	2021	3	30189	30486	Colorado Springs, CO	11189	14207	COS	PHX	551	261	120.59	WN	0.8219	124.41	F9	0.1164	67.59	20210319114207COSPHX						
38	Tabelle	2021	3	30189	30852	Colorado Springs, CO	11189	10821	COS	BWI	1503	65	251.59	WN	0.5386	238.9	WN	0.5386	238.9	2021031910821COSBWI						
39	Tabelle	2021	3	30189	30852	Colorado Springs, CO	11189	11278	COS	DCA	1503	84	278.14	AA	0.5724	287.79	WN	0.1886	199.98	2021031911278COSDCA						
40	Tabelle	2021	3	30189	30852	Colorado Springs, CO	11189	12064	COS	IAH	1503	63	301.82	UA	0.6045	342.52	WN	0.2763	200.22	2021031912064COSIAH						
41	Tabelle	2021	3	30189	32575	Colorado Springs, CO	11189	10800	COS	BUR	833	33	183.97	WN	0.6558	159.2	WN	0.6558	159.2	2021031910800COSBUR						
42	Tabelle	2021	3	30189	32575	Colorado Springs, CO	11189	12892	COS	LAX	833	109	203.47	UA	0.6444	221.23	WN	0.2251	153.75	2021031912892COSLAX						
43	Tabelle	2021	3	30189	32575	Colorado Springs, CO	11189	12954	COS	LGB	833	24	143.75	WN	0.8571	120.82	WN	0.8571	120.82	2021031912954COSLGB						
44	Tabelle	2021	3	30189	32575	Colorado Springs, CO	11189	13891	COS	ORD	833	46	188.22	WN	0.5619	171.16	WN	0.5619	171.16	2021031913891COSORD						
45	Tabelle	2021	3	30189	32575	Colorado Springs, CO	11189	14908	COS	SNA	833	48	200.83	WN	0.5666	170.91	WN	0.5666	170.91	2021031914908COSSNA						
46	Tabelle	2021	3	30194	30198	Dallas/Fort Worth, TX	11194	14122	DAL	PIT	1067	65	209.05	WN	0.9121	206.89	WN	0.9121	206.89	20210312594122DALPIT						
47	Tabelle	2021	3	30194	30198	Dallas/Fort Worth, TX	11194	11298	DAL	DFW	1067	432	238.32	AA	0.9359	240.38	NK	0.0138	95.38	2021031259412298DALDFW						
48	Tabelle	2021	3	30194	30253	Dallas/Fort Worth, TX	11194	11292	DAL	DEN	651	1033	134.1	WN	0.9985	134.14	WN	0.9985	134.14	2021031259412292DALDEN						

2. Uploaded the above to ADLS storage and added to the source and checked schema for the data.

Source settings

Source options

Projection

Optimize

Inspect

Data preview

Number of columns **Total** 23

Order	Column	Type
1	tbl	abc string
2	Year	abc string
3	quarter	abc string
4	citymarketid_1	abc string
5	citymarketid_2	abc string
6	city1	abc string
7	city2	abc string

- Applied filter to see only flights Starting from Dallas Texas (applied filter for column citymarketid\_1 == 30194)

Reference: 1  
Columns: 23 total

FilterflightsfromDallasTX  
Columns: 23 total

Filter settings | Optimize | Inspect | Data preview

Output stream name \* FilterflightsfromDallasTX

Description Filtering rows using expressions on columns 'citymarketid\_1'

Incoming stream \* Sourcedata

Filter on \* citymarketid\_1 == 30194

- Used Select command and narrowed down the columns and selected only columns Year, city 1, city 2, airport 1, airport2, passengers, fare,.

Select settings | Optimize | Inspect | Data preview

☒ Skip duplicate output columns ⓘ

Input columns \* ☐ Auto mapping ⓘ    7 mappings: 16 column(s) from the inputs left unmapped

<input type="checkbox"/>	FilterflightsfromDallasTX's column		Name as	
<input type="checkbox"/>	abc Year	→	Year	<input type="button" value="+"/> <input type="button" value="Delete"/>
<input type="checkbox"/>	abc city1	→	city1	<input type="button" value="+"/> <input type="button" value="Delete"/>
<input type="checkbox"/>	abc city2	→	city2	<input type="button" value="+"/> <input type="button" value="Delete"/>
<input type="checkbox"/>	abc airport_1	→	airport_1	<input type="button" value="+"/> <input type="button" value="Delete"/>
<input type="checkbox"/>	abc airport_2	→	airport_2	<input type="button" value="+"/> <input type="button" value="Delete"/>
<input type="checkbox"/>	abc passengers	→	passengers	<input type="button" value="+"/> <input type="button" value="Delete"/>
<input type="checkbox"/>	abc fare	→	fare	<input type="button" value="+"/> <input type="button" value="Delete"/>

5. Sorted the year in descending order.

Sort settings   Optimize   Inspect   Data preview

Output stream name \*  [Learn more](#)

Description  [Reset](#)

Incoming stream \*

Options \* ☐ Case insensitive  
☐ Sort only within partition

Sort conditions \* 

Selectfewcoulmns's column	Order	Nulls first
<input type="text" value="abc Year"/>	<input type="text" value="Descending"/>	<input checked="" type="checkbox"/>

6. Extracted to our ADLS output target Sink folder.

Sink   Settings   Errors   Mapping   Optimize   Inspect   Data preview

Output stream name \*  [Learn more](#)

Description  [Reset](#)

Incoming stream \*

Sink type \* 

Dataset

Inline

Cache

Dataset \*  [Test connection](#) [Open](#) [New](#)

Skip line count

Options ☒ Allow schema drift

## 7. Executed the Dataflow into a pipeline and got the below desired result.

✓ Validate > Debug > Add trigger Data flow debug

Data flow

Data flow1

Parameters Variables Settings **Output**

Pipeline run ID: c85add6c-2e8a-441a-bd17-5106a9048604 Pipeline status: Succeeded

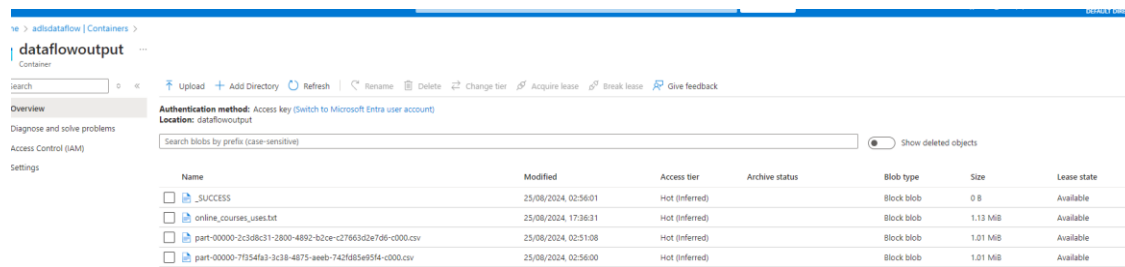
All status

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties	Activity run ID	Log
Data flow1	Succeeded	Data flow	8/25/2024, 6:50:56 AM	52s	AutoResolveIntegration		c1da3ac3-4d77-4619-8084-956a75382d1	

	A	B	C	D	E	F	G	H
1	Year	city1	city2	airport_1	airport_2	passenger	fare	
2	2024	Dallas/Fort Worth	Pittsburgh	DAL	PIT	117	224.91	
3	2024	Dallas/Fort Worth	Pittsburgh	DFW	PIT	411	309.38	
4	2024	Dallas/Fort Worth	Denver, CO	DAL	DEN	1262	163.48	
5	2024	Dallas/Fort Worth	Denver, CO	DFW	DEN	1718	201.24	
6	2024	Dallas/Fort Worth	Tucson, AZ	DAL	TUS	35	234.66	
7	2024	Dallas/Fort Worth	Tucson, AZ	DFW	TUS	265	312.99	
8	2024	Dallas/Fort Worth	Phoenix, AZ	DAL	PHX	1119	207.6	
9	2024	Dallas/Fort Worth	Phoenix, AZ	DFW	PHX	1716	260.34	
10	2024	Dallas/Fort Worth	Hartford, CT	DAL	BDL	33	277.13	
11	2024	Dallas/Fort Worth	Hartford, CT	DFW	BDL	203	362.81	
12	2024	Dallas/Fort Worth	Seattle, WA	DAL	SEA	225	234.54	
13	2024	Dallas/Fort Worth	Seattle, WA	DFW	SEA	1185	282.6	
14	2024	Dallas/Fort Worth	El Paso, TX	DAL	ELP	433	187.17	
15	2024	Dallas/Fort Worth	El Paso, TX	DFW	ELP	289	254.43	
16	2024	Dallas/Fort Worth	Nashville, TN	DAL	BNA	636	213.38	
17	2024	Dallas/Fort Worth	Nashville, TN	DFW	BNA	712	262.1	
18	2024	Dallas/Fort Worth	Washington, DC	DAL	BWI	586	213.97	
19	2024	Dallas/Fort Worth	Washington, DC	DAL	DCA	529	192.79	
20	2024	Dallas/Fort Worth	Washington, DC	DAL	IAD	7	242.22	
21	2024	Dallas/Fort Worth	Washington, DC	DFW	BWI	648	199.27	
22	2024	Dallas/Fort Worth	Washington, DC	DFW	DCA	1003	310.28	
23	2024	Dallas/Fort Worth	Washington, DC	DFW	IAD	584	255.64	
24	2024	Dallas/Fort Worth	St. Louis, MO	DAL	STL	589	213.63	
25	2024	Dallas/Fort Worth	St. Louis, MO	DFW	STL	400	267.45	
26	2024	Dallas/Fort Worth	Jacksonville, FL	DAL	JAX	51	238.98	
27	2024	Dallas/Fort Worth	Jacksonville, FL	DFW	JAX	475	291.64	
28	2024	Dallas/Fort Worth	Detroit, MI	DAL	DTW	42	244.15	
29	2024	Dallas/Fort Worth	Detroit, MI	DFW	DTW	1151	246.73	
30	2024	Dallas/Fort Worth	Des Moines, IA	DAL	DSM	14	207.58	
31	2024	Dallas/Fort Worth	Des Moines, IA	DFW	DSM	219	274.19	
32	2024	Dallas/Fort Worth	Houston, TX	DAL	HOU	928	187.27	
33	2024	Dallas/Fort Worth	Houston, TX	DAL	IAH	116	163.95	
34	2024	Dallas/Fort Worth	Houston, TX	DFW	HOU	150	228.3	
35	2024	Dallas/Fort Worth	Houston, TX	DFW	IAH	746	217.96	

8. See below the output has been to stored in the sink (Target) location:

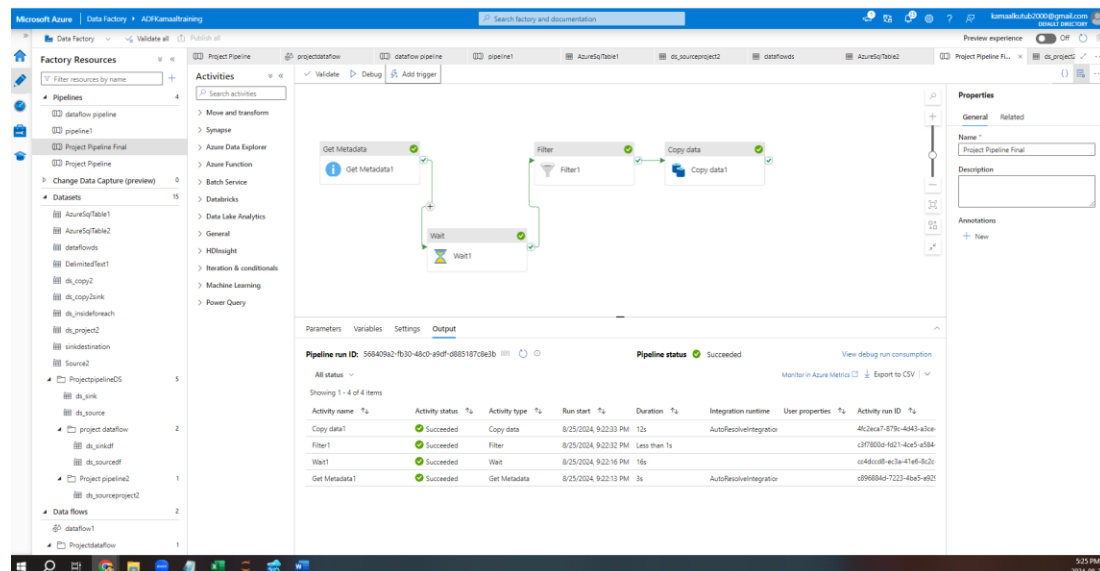


The screenshot shows the 'dataflowoutput' container in the Azure Data Factory portal. It displays a list of blobs stored in the sink. The table below represents the data shown in the screenshot.

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
_SUCCESS	25/08/2024, 02:56:01	Hot (inferred)		Block blob	0 B	Available
online_courses_users.txt	25/08/2024, 17:36:31	Hot (inferred)		Block blob	1.13 MiB	Available
part-00000-2cd8c31-2800-4892-b2ce-c27663d2e7d6-c000.csv	25/08/2024, 02:51:08	Hot (inferred)		Block blob	1.01 MiB	Available
part-00000-7f954fa3-3c18-4875-ae6b-742f8b5e95f4-c000.csv	25/08/2024, 02:56:00	Hot (inferred)		Block blob	1.01 MiB	Available

**ADF Pipeline Project:** Creating a Pipeline to filter a file from a set of files and copy to the desired destination.

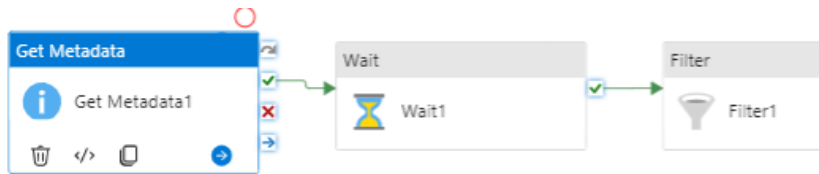
From the Container of Adls wanted to filter single file and copied it to the desired storage location.



The screenshot shows the 'Project Pipeline Final' pipeline in the Azure Data Factory portal. The pipeline consists of four activities: 'Get Metadata1', 'Filter1', 'Copy data1', and 'Wait1'. The pipeline status is 'Succeeded'. The output table below shows the execution details for each activity.

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties	Activity run ID
Copy data1	Succeeded	Copy data	8/25/2024, 9:22:33 PM	12s	AutoResolveIntegrator		4f2eac1-8776-4943-a30a-c8f7800a-fd21-4ae5-d584
Filter1	Succeeded	Filter	8/25/2024, 9:22:32 PM	Less than 1s			cd4d0c8-e43a-41af-8c2c-c8968046-7223-4ba3-e92c
Wait1	Succeeded	Wait	8/25/2024, 9:22:16 PM	16s			
Get Metadata1	Succeeded	Get Metadata	8/25/2024, 9:22:13 PM	3s	AutoResolveIntegrator		

1. Taken the meta data to call for the Folder (container) where a number of files are located.



General Settings User properties

Dataset \*

Field list \*

☐ Argument

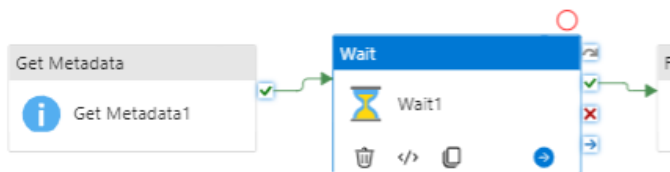
☐ Child items

Filter by last modified ⓘ

Start time (UTC)

End time (UTC)

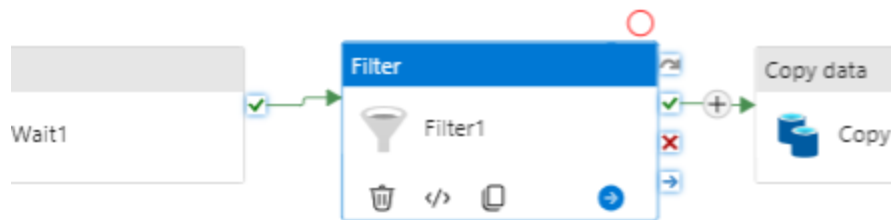
2. Used wait for 15 seconds to give a wait time



General Settings User properties

Wait time in seconds \*

3. Filtered the desired file – Online Courses from a number of files present in the container.



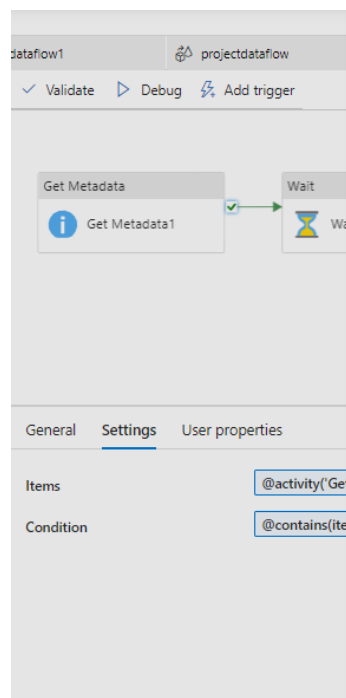
General Settings User properties

Items

@activity('Get Metadata1').output.chi...

Condition

@contains(item().name, 'online\_cour...



#### Pipeline expression builder

Add dynamic content below using any combination of [expressions](#), [functions](#) and [system variables](#).

@activity('Get Metadata1').output.childItems

[Clear contents](#)

Activity outputs Parameters System variables Functions Variables

Search

Get Metadata1

Get Metadata1 activity output

Get Metadata1

Get Metadata1 pipeline return value

Get Metadata1 childItems

List of subfolders and files in the given folder

Get Metadata1 exists

Whether a file, folder, or table exists

OK

Cancel

dataflow1 projectdataflow

✓ Validate ▶ Debug ⚡ Add trigger

Get Metadata  
Get Metadata1

Wait  
Wait1

General Settings User properties

Items @activity('Get Metadata1')

Condition @contains(item().name, 'online\_courses\_uses.csv')

### Pipeline expression builder

Add dynamic content below using any combination of [expressions](#), [functions](#) and [system variables](#).

```
@contains(item().name, 'online_courses_uses.csv')
```

[Clear contents](#)

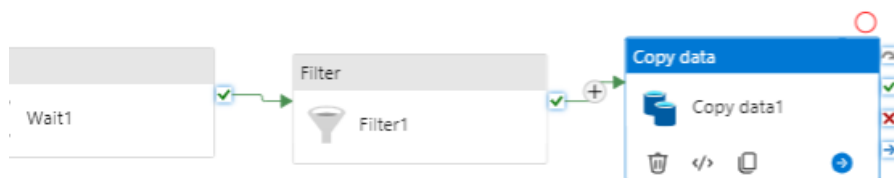
Filter iterator Activity outputs Parameters System variables Functions \

Search

Filter1  
Current item

OK Cancel

4. Copied the online courses file to desired location of storage in ADLS.



General Source Sink Mapping Settings User properties

Name \* Copy data1 [Learn more](#)

Description

Activity state ☒ Activated ☐ Deactivated

Timeout 0.12:00:00

Retry 0



✓ Validate ▶ Debug ⚙️ Add trigger

```
graph LR; A[Get Metadata1] --> B[Wait1]; B --> C[Filter1]; C --> D[Copy data1];
```

Parameters Variables Settings **Output**

**Pipeline run ID:** 65404576-ca28-4c2a-b7af-59af2eaf5402 🔄 🔍 📄 **Pipeline status:** ✓ Succeeded [View debug run consumption](#)

All status ▼ [Monitor in Azure Metrics](#) [Export to CSV](#) ▼

Showing 1 - 4 of 4 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties	Activity run ID
Copy data1	<span>✓</span> Succeeded	Copy data	8/25/2024, 9:36:19 PM	14s	AutoResolveIntegration		d34ef1b8-8bce-46a7-a9f7-
Filter1	<span>✓</span> Succeeded	Filter	8/25/2024, 9:36:18 PM	Less than 1s			5cd5db43-f126-4610-86f4
Wait1	<span>✓</span> Succeeded	Wait	8/25/2024, 9:36:02 PM	16s			d75ea910-799c-4eb8-a70-
Get Metadata1	<span>✓</span> Succeeded	Get Metadata	8/25/2024, 9:35:59 PM	3s	AutoResolveIntegration		ecf792a5-e705-48f8-a5df-

6. See below confirmation the copy activity executed the file to destination.

Microsoft Azure 🔍 Search resources, services, and docs (5 +) Copilot 🔍 🔔 ⚙️ 📄 👤 kamaufutub2000@gm... DEFAULT DIRECTORY

🏠 > adsdatalog | Containers >

### dataflowoutput

Container

Search  📁 Upload ➕ Add Directory 🔄 Refresh 🏷️ Rename 🗑️ Delete 🔄 Change tier 🔑 Acquire lease 🔑 Break lease 🗨️ Give feedback

**Overview** 🔍 🔧 Diagnose and solve problems 🔒 Access Control (IAM) ⚙️ Settings

**Authentication method:** Access key [\(Switch to Microsoft Entra user account\)](#)  
**Location:** dataflowoutput

Search blobs by prefix (case-sensitive)  🔍 Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/> _SUCCESS	25/08/2024, 02:56:01	Hot (inferred)		Block blob	0 B	Available ***
<input type="checkbox"/> online_courses_usgs.txt	25/08/2024, 17:36:31	Hot (inferred)		Block blob	1.13 MiB	Available ***