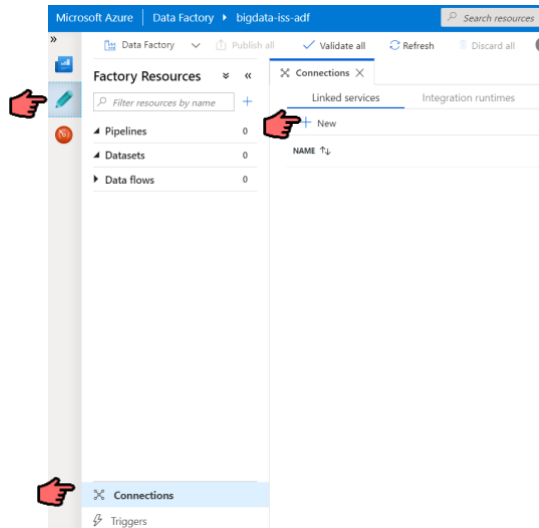


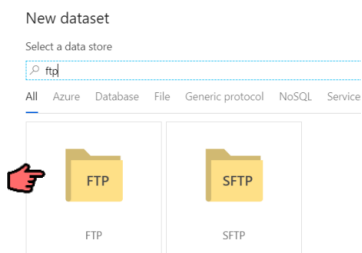
Via FTP source to Blob sink

Create source connection service

1. Go to Azure data factory. Go to Author→**Connections**. Click +New



2. Select FTP



3. Fill in 138.75.113.64 (subject to changes) in Host field.
Select **Disable SSL** transmission
Select **Basic** authentication type (Anonymous should also work)

New linked service (FTP)

Name *

FtpServer1

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Host *

138.75.113.64

Port

21

Secure Transmission

☐ Enable SSL ☒ Disable SSL

Authentication type *

Basic

User name *

bigdata

- Key in user name: *bigdata*
Password: *BigData_12345*
Test connection to ensure its successful.
Click **Create**

User name *

bigdata

Password *

Annotations

+ New

Create Back

Connection successful

Test connection Cancel

Create destination linked service

- Add another new linked service. Select **Azure Blob Storage**

Data store Compute

Search

All Azure Database File Generic protocol NoSQL

Amazon Marketplace Web Service

Amazon Redshift

Amazon S3

Apache Impala

Azure Blob Storage

Azure Cosmos DB (MongoDB API)

- Set storage account name to **bigdataisspahdisstorage** (or other storage of your choice).
Test connection. If successful, click on **Create**.

New linked service (Azure Blob Storage)

Name *
AzureBlobStorage1

Description

Connect via integration runtime *
AutoResolveIntegrationRuntime

Authentication method
Account key

Connection string Azure Key Vault

Account selection method
☒ From Azure subscription ☐ Enter manually

Azure subscription
Select all

Storage account name *
bigdataisspahdistorage

Additional connection properties
+ New

Create source dataset

7. Create new dataset

Pipelines 1
pipeline1

Datasets 0

Data flows

Search activities

Move & transform

Azure Data Explorer

New dataset

New folder

Data Lake Analytics

Databricks

8. Select FTP. Next select DelimitedText

New dataset

Select a data store

ftp

All Azure Database File Generic protocol NoSQL Service

FTP SFTP

Select format

Choose the format type of your data

Parquet DelimitedText Json

Avro ORC Binary

9. Set linked service to **FtpServer1** (earlier created).
Set File path to navigate to ftp->files->(any one file).
Check **First row as header**.
Click Ok.

Set properties

Name
DelimitedText1

Linked service *
FtpServer1
[Edit connection](#)

File path
files / BSE_BOM539917_sample.csv [Browse](#) ▼

First row as header ☒

Import schema
☒ From connection/store ☐ From sample file ☐ None

Create destination dataset

10. Add another new dataset. Select **Azure Blob Storage**, followed by **DelimitedText**

11. Set linked service to **AzureBlobStorage1** (earlier created).

Set File path to navigate to **ingestion** container

Key in output file name

Check **First row as header**.

Set import scheme to **None**.

Click Ok.

Set properties

Name
Dest_DelimitedText1

Linked service *
AzureBlobStorage1
[Edit connection](#)

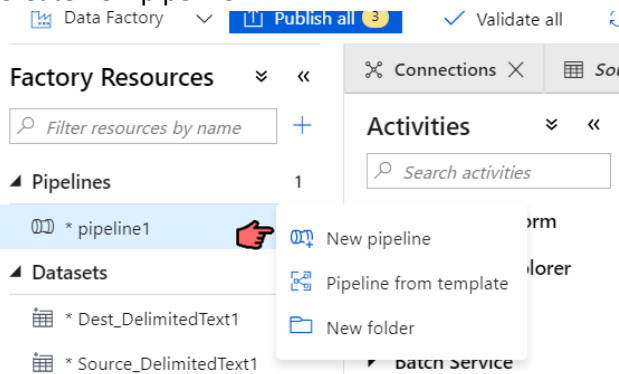
File path
ingestion / Directory ftpoutput [Browse](#) ▼

First row as header ☒

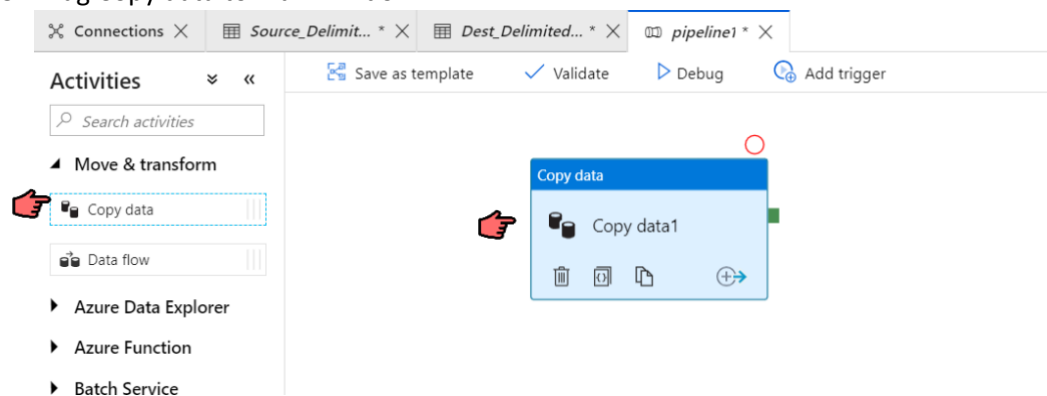
Import schema
☐ From connection/store ☐ From sample ☒ None

Create dataflow pipeline

12. Create new pipeline.



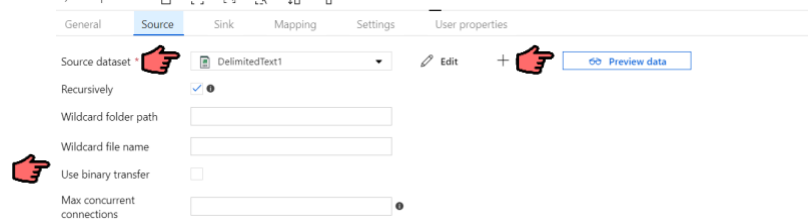
13. Drag Copy data to main window




14. Set Source to earlier source dataset created.

Click on Preview data to see how the data format would appear when ingested.

Uncheck **Use binary transfer**






Data Preview 


Linked service: FtpServer1
Object: BSE_BOM539917_sample.csv


date	Open	High	Low	Close	WAP	No.of.Shares	No.of.Trades	Total.Turnover	Deliverable.Quantity	X.Deli.
2017-02-21	11.74	12	11.65	11.94	11.82	106699	240	1261289	106699	100
2017-02-22	12.45	12.53	11.9	12.06	12.39	352673	445	4370446	234847	66.59
2017-02-23	12.15	12.51	11.46	12.25	12.06	421748	708	5085943	201859	47.86
2017-02-27	12.25	12.39	11.75	11.82	12.07	95021	189	1146935	76519	80.53
2017-02-28	11.8	12.2	11.66	11.77	11.82	146428	235	1731418	93590	63.92
2017-03-01	11.52	12.24	11.52	12.14	11.94	133832	202	1598152	79794	59.62
2017-03-02	12.21	12.73	11.93	11.98	12.25	146380	237	1793472	101823	69.56
2017-03-03	12.1	12.16	11.87	11.92	11.95	36942	105	441440	28273	76.53
2017-03-06	11.9	12.33	11.9	12.16	12.19	106479	225	1298314	64491	60.57

15. Set sink to Destination data set earlier created.
Rename file extension to appropriate form.


General Source **Sink** Mapping Settings User properties

Sink dataset *  DelimitedText2  Edit  New

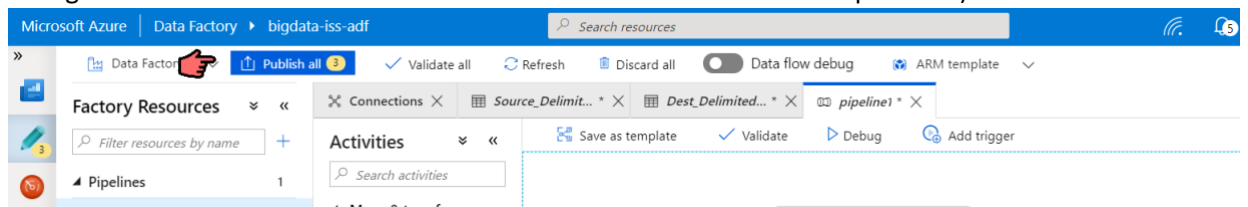
Copy behavior None 

Max concurrent connections 

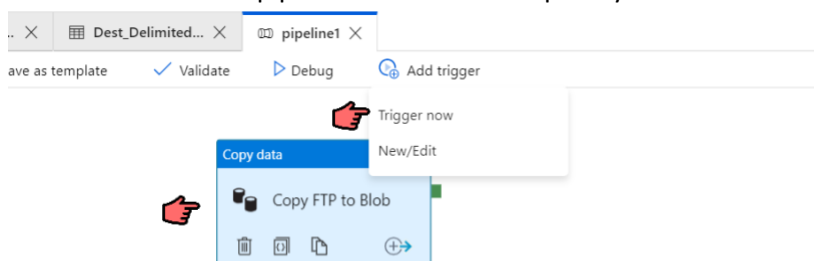
Quote all text ☒

File extension  .CSV

16. Publish **all changes**. Wait for changes to deployed successfully. (This is to **effect** all the changes we have made from our local workstation over to Azure cloud platform).



17. Select pipeline earlier create. Click **Add trigger** → **trigger now**.
Click Finish. Wait for pipeline to execute completely.



18. Go to specified storage account container to affirm file is successfully ingested.