

we create

a datafactory

a linked service pointing to http

a linked service pointing to datalake

<https://trendytechretailsa.blob.core.windows.net/retail-raw/products.csv>

https://trendytechretailsa.blob.core.windows.net/retail-raw/order_items.csv

<https://trendytechretailsa.blob.core.windows.net/retail-raw/orders.csv>

https://files.cdn.thinkific.com/file_uploads/349536/attachments/e8e/2f1/eca/departmen

ts.csv

https://files.cdn.thinkific.com/file_uploads/349536/attachments/5f1/f7c/44a/customers.

csv

https://files.cdn.thinkific.com/file_uploads/349536/attachments/8ad/0ec/cdf/categories.

csv

if you want to ingest all of the above 6 files

2 http linked services - source

<https://trendytechretailsa.blob.core.windows.net>

<https://files.cdn.thinkific.com>

1 linked service pointing to datalake

12 datasets

6 for the source - 6 different relative urls

6 for the target - 6 different filenames

you would require 6 different copy activities

6 different pipelines

To avoid creating so many components we would parameterize the pipeline

pipeline

copy activity

Source Linked Service - baseUrl

Datasets - relativeURL

Target Linked Service - fileName

pipeline

copy data activity

Data Set -

Linked Service - we have parameterized (baseUrl)

=====

2 files

=====

orders.csv

order_items.csv

orders.csv will be coming in your blob storage

order_items.csv will be uploaded in amazon s3

as soon as orders.csv arrives in the blob storage I want to ingest both
order.csv from

blob and order_items.csv from s3 to adls gen2..

my ingestion pipeline should trigger on arrival of orders.csv file.

I want to divide these orders in 3 parts...

in my datalake

high_value_order > 500

low_value_order < = 500

erroneous_order (all the remaining orders)

in sql table we need to ingest the high value orders for reporting purpose..

resource group

storage account

data lake

data factory

amazon s3 bucket

Key vault

=====

orders.csv

blob storage - azure

adls - azure

blob

s3

adls

key vault

=====

ingest order items - s3

ingest the orders data - blob

Process part

=====

high value

low value

erroneous

output - 3 different folders..

to take the high value orders and populate this in our azure sql database for the

reporting team.

a linked service to point to adls - done

a linked service to point to azure sql database

a dataset which refers to high value order in my datalake - done

a dataset which will refer to the table in sql database

pipeline

scheduled trigger - you can give a future date

tumbling window - deal with slices of data

we can run it for a past interval also

storage event - based on creation or deletion of files

custom event

===

just like we have chained up the pipelines.. we can even chain up the triggers...

only supported for tumbling window trigger

=====

scheduled trigger

one trigger can invoke multiple pipelines

many triggers can be attached to one pipeline

many to many

=====

tumbling window triggers

we cannot have a many to many relationship here

=====

```
CREATE TABLE premium_orders (  
  order_id INT NOT NULL,  
  order_date VARCHAR(45) NOT NULL,  
  order_customer_id INT NOT NULL,  
  order_status VARCHAR(45) NOT NULL,  
  order_amount float,  
  PRIMARY KEY (order_id)  
);
```

=====

=====

2 files

=====

orders.csv

order_items.csv

orders.csv will be coming in your blob storage

order_items.csv will be uploaded in amazon s3

as soon as orders.csv arrives in the blob storage I want to ingest both
order.csv from blob and

order_items.csv from s3 to adls gen2..

my ingestion pipeline should trigger on arrival of orders.csv file.

I want to divide these orders in 3 parts...

in my datalake

high_value_order > 500

low_value_order < = 500

erroneous_order (all the remaining orders)

in sql table we need to ingest the high value orders for reporting purpose..

resource group

storage account

data lake

data factory

amazon s3 bucket

Key vault

=====

orders.csv

blob storage - azure

adls - azure

blob

s3

adls

key vault

=====

ingest order items - s3

ingest the orders data - blob

Process part

=====

high value

low value

erroneous

output - 3 different folders..

to take the high value orders and populate this in our azure sql database for the reporting team.

a linked service to point to adls - done

a linked service to point to azure sql database

a dataset which refers to high value order in my datalake - done

a dataset which will refer to the table in sql database

pipeline

scheduled trigger - you can give a future date

tumbling window - deal with slices of data

we can run it for a past interval also

storage event - based on creation or deletion of files

custom event

===

just like we have chained up the pipelines.. we can even chain up the triggers...

only supported for tumbling window trigger

=====

scheduled trigger

one trigger can invoke multiple pipelines

many triggers can be attached to one pipeline

many to many

=====

tumbling window triggers

we cannot have a many to many relationship here

=====

```
CREATE TABLE premium_orders (  
  order_id INT NOT NULL,  
  order_date VARCHAR(45) NOT NULL,  
  order_customer_id INT NOT NULL,  
  order_status VARCHAR(45) NOT NULL,  
  order_amount float,  
  PRIMARY KEY (order_id)  
);
```

=====

