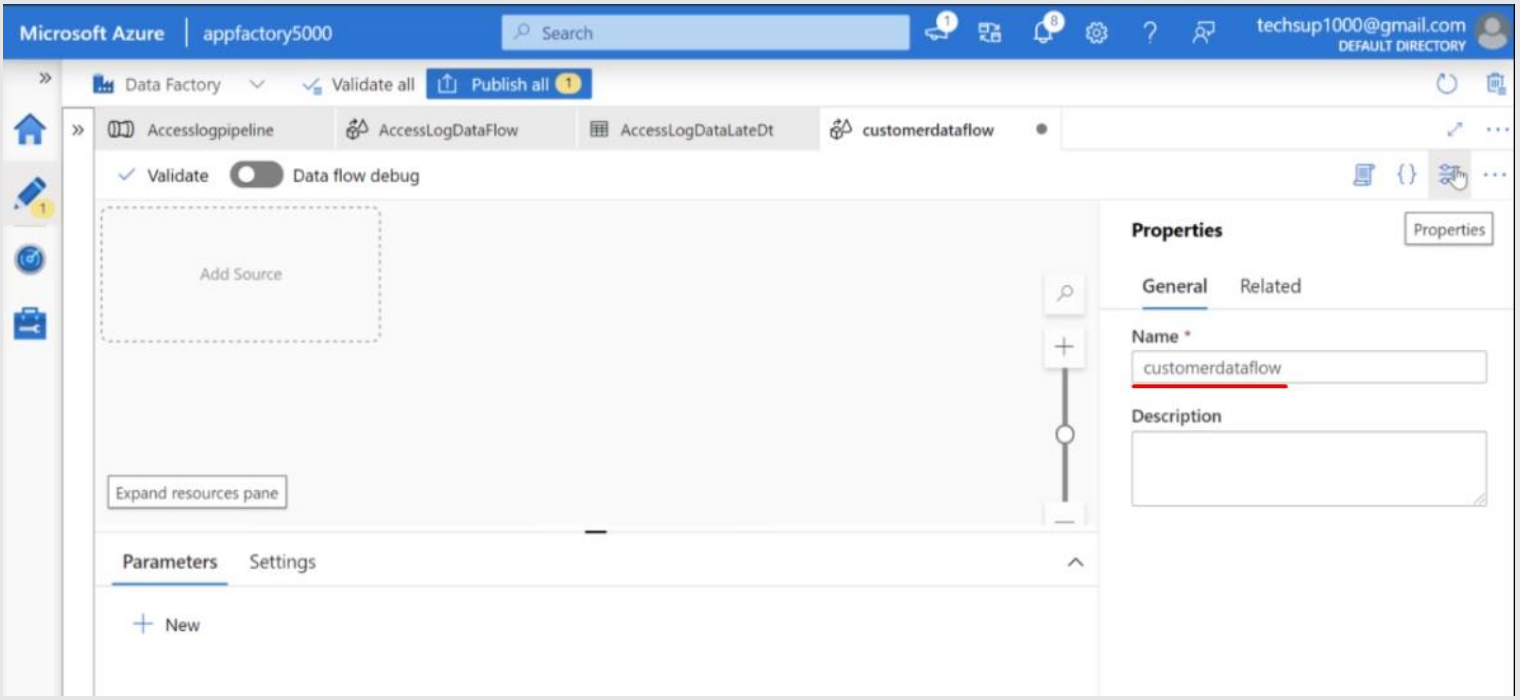
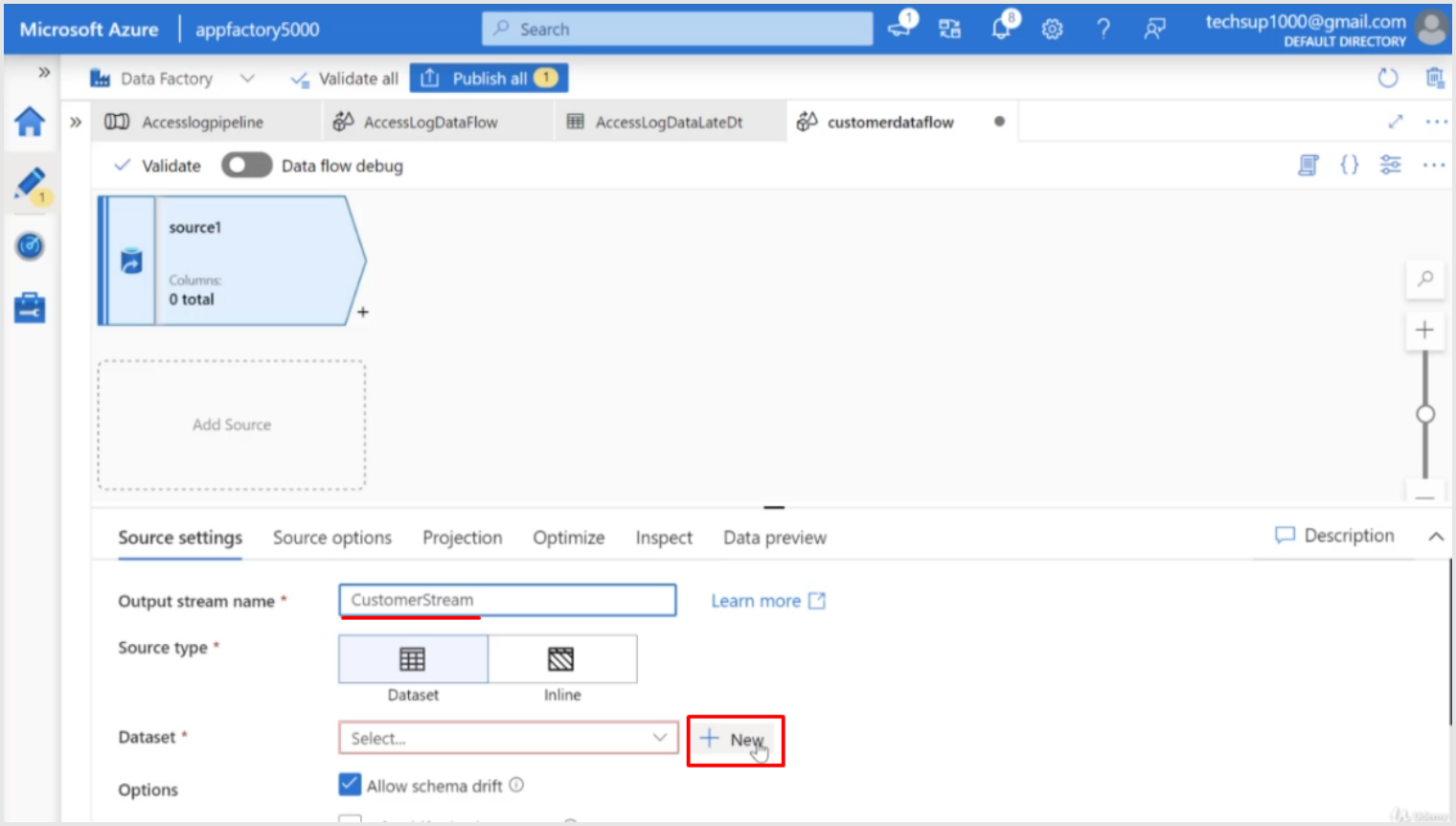
Azure Data Factory Cargar un archivo JSON Object a un Dedicated SQL Pool

Se busca procesar el siguiente archivo **JSON Object**. Básicamente es el mismo archivo que el punto 14, pero se le agrego el elemento “details” que vendría a ser del tipo **‘Object’**. En Spark vendría a ser del tipo **‘Struct’**.

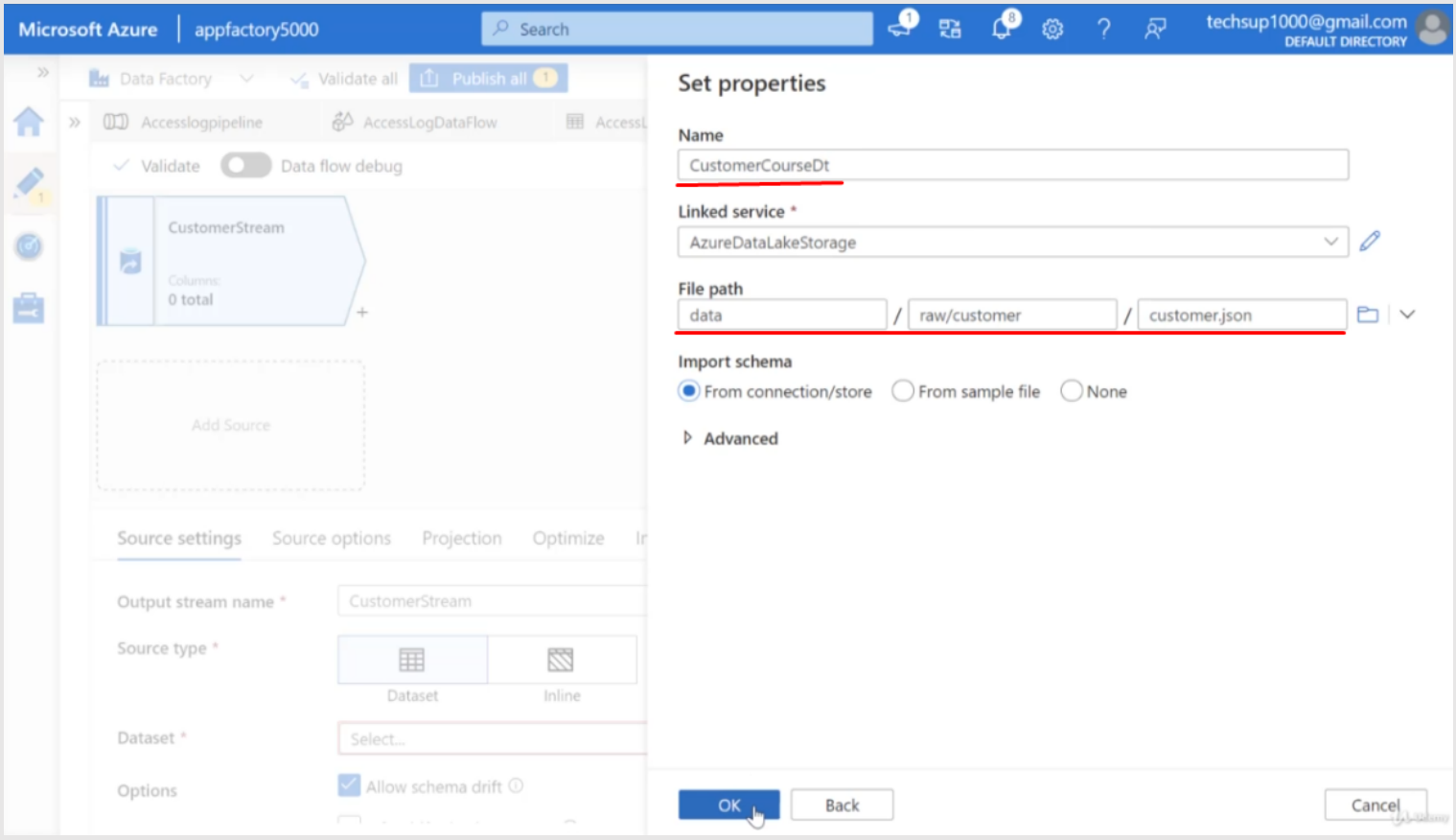
```json
[
{
"customerid":1,
"customername":"UserA",
"registered":true,
"courses":["AZ-900","AZ-500","AZ-303"],
"details" :
    {
    "mobile":"111-1112",
    "city":"CityA"
    }
},
{
"customerid":2,
"customername":"UserB",
"registered":true,
"courses":["AZ-104","AZ-500","DP-200"],
"details" :
    {
    "mobile":"333-1112",
    "city":"CityB"
    }
}]
```

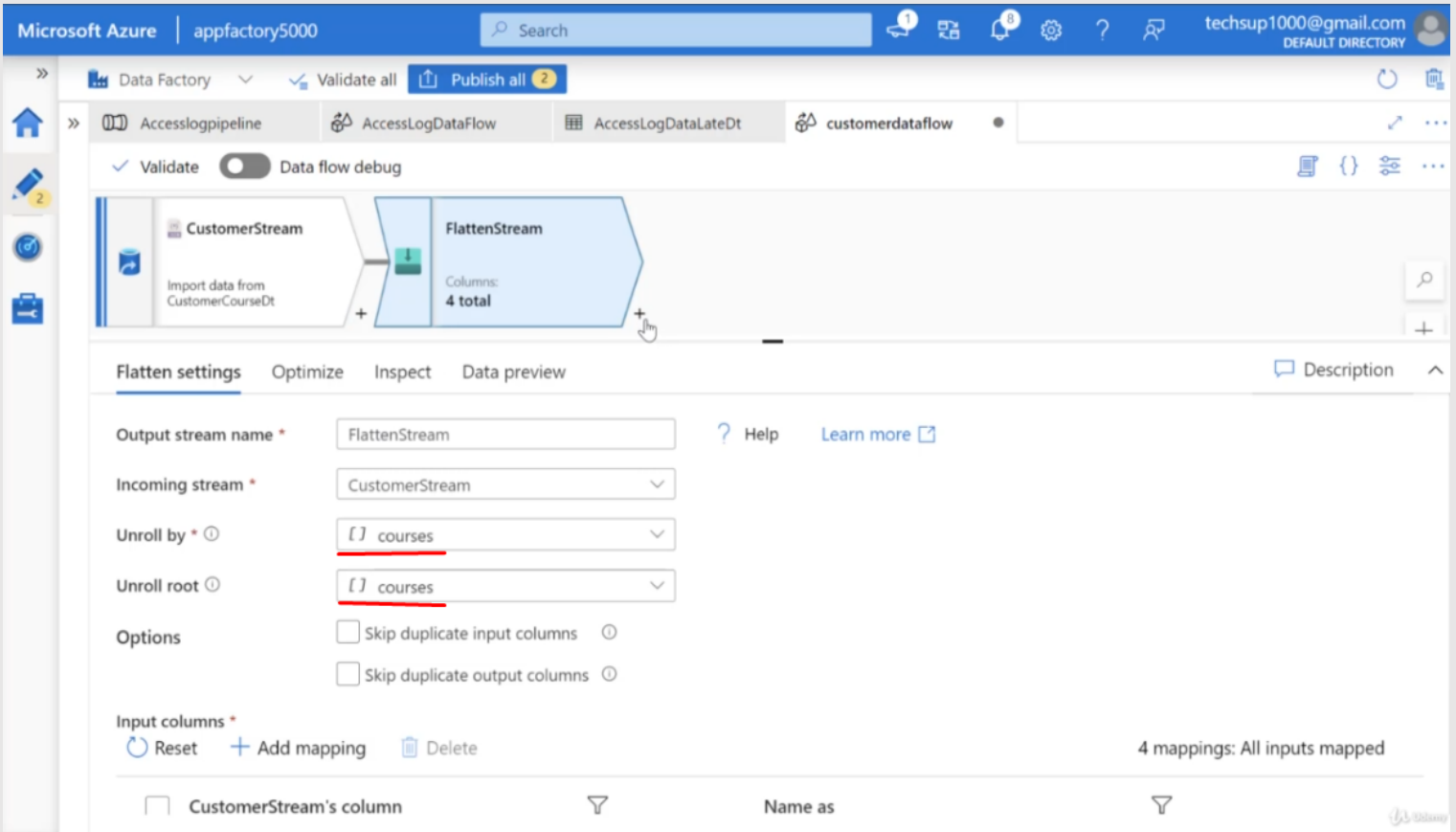
Comenzamos creando nuestro Dataflow

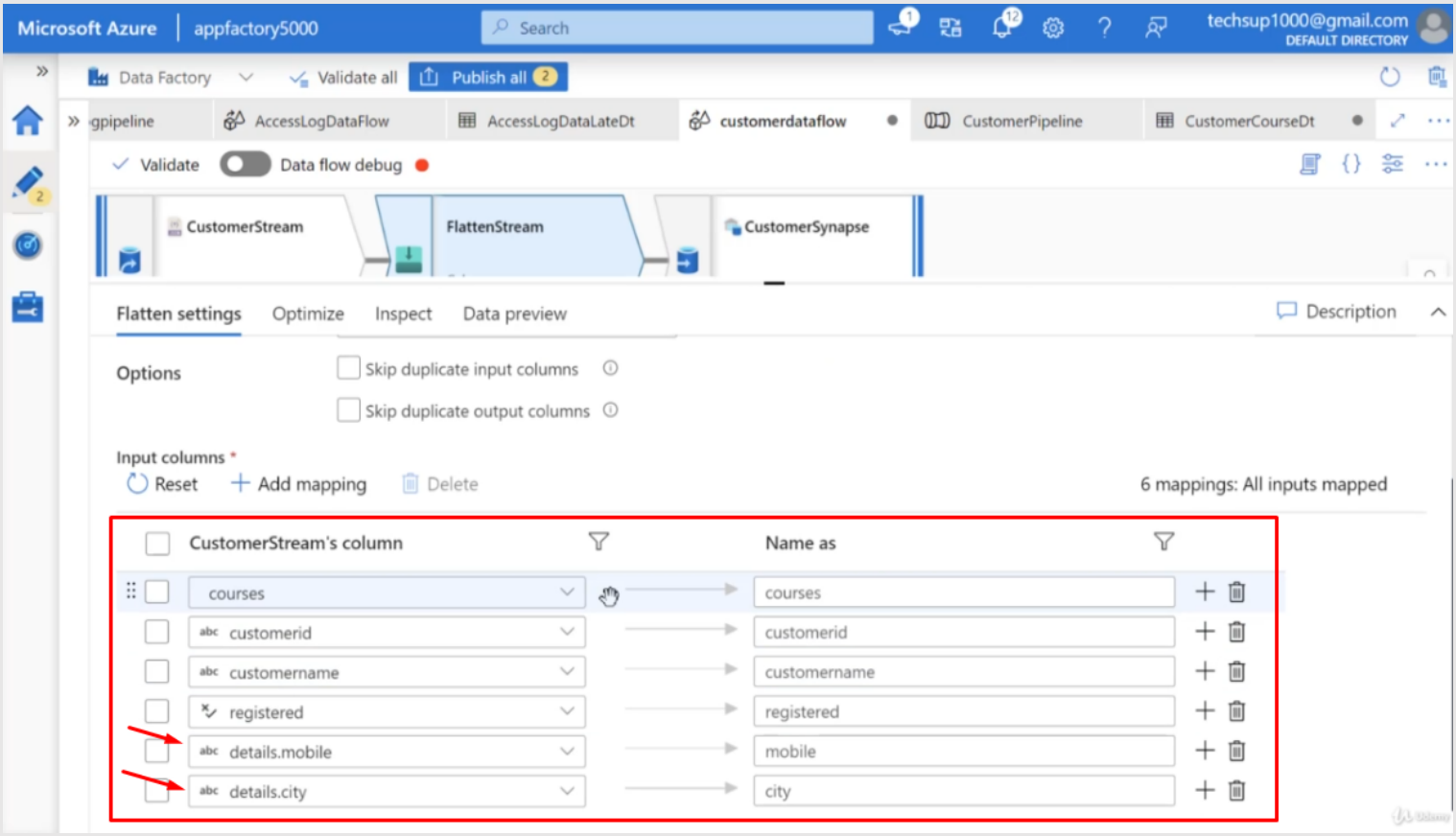


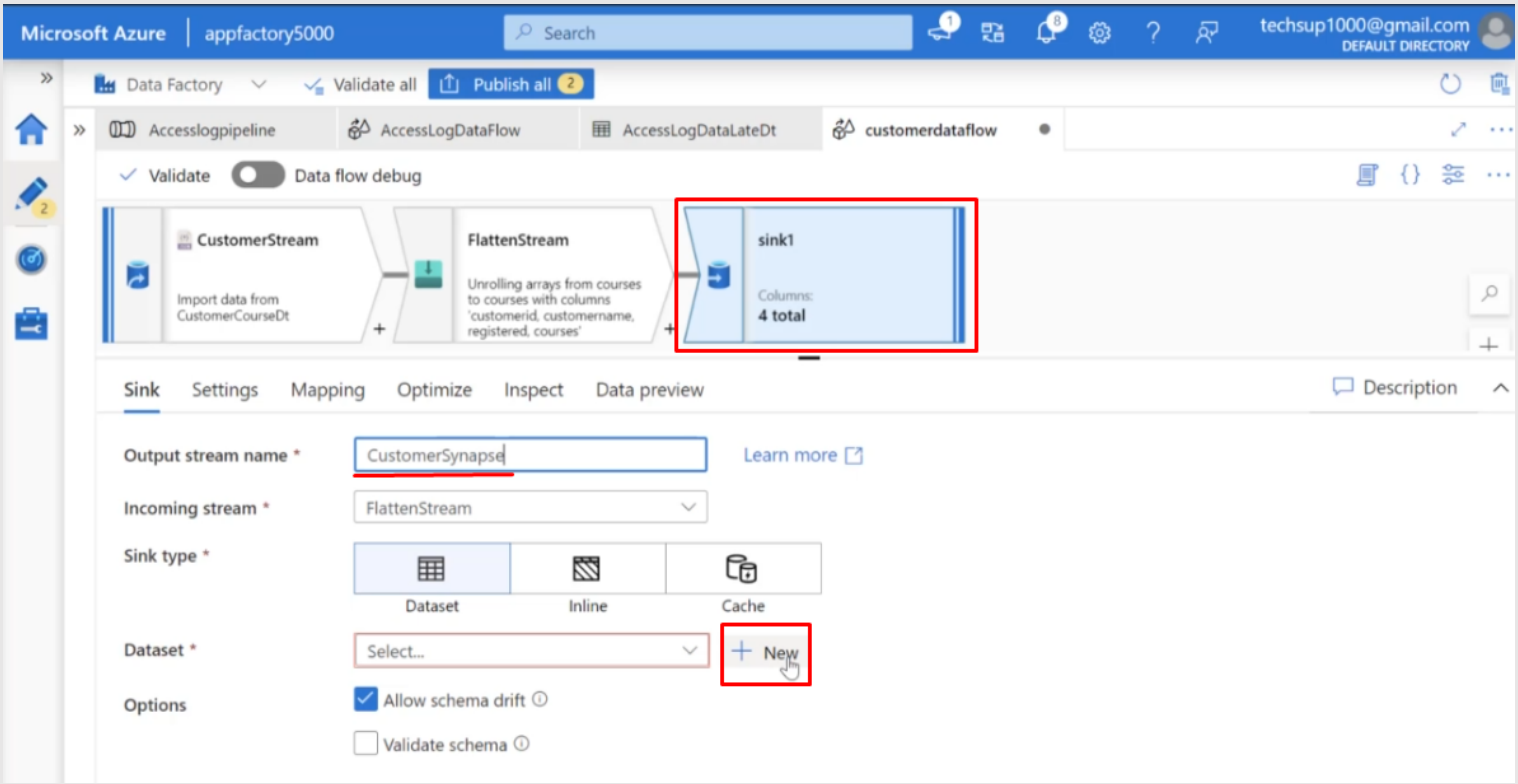


Cargaremos el archivo desde Azure Data Lake Storage y será un archivo JSON:

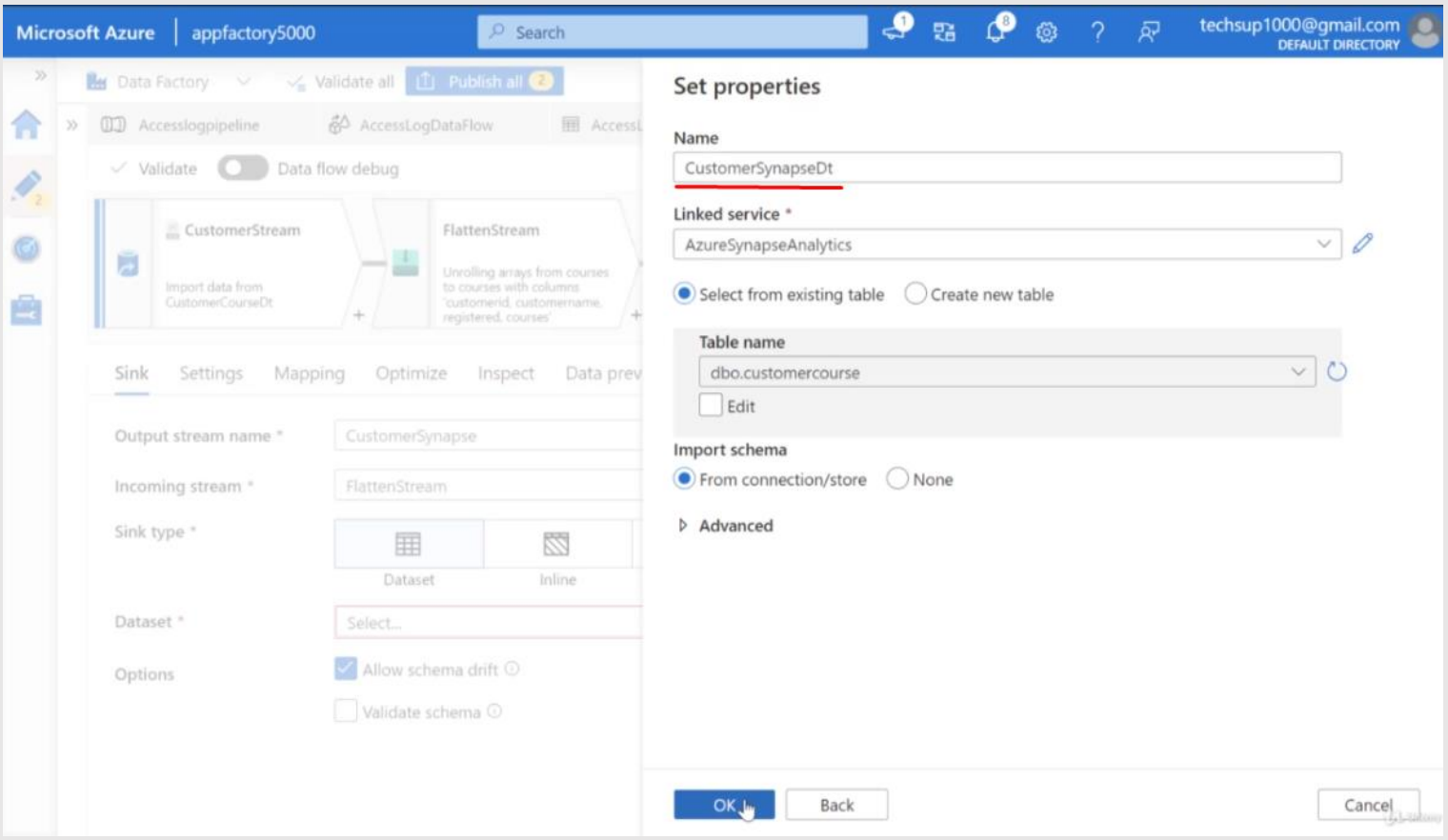




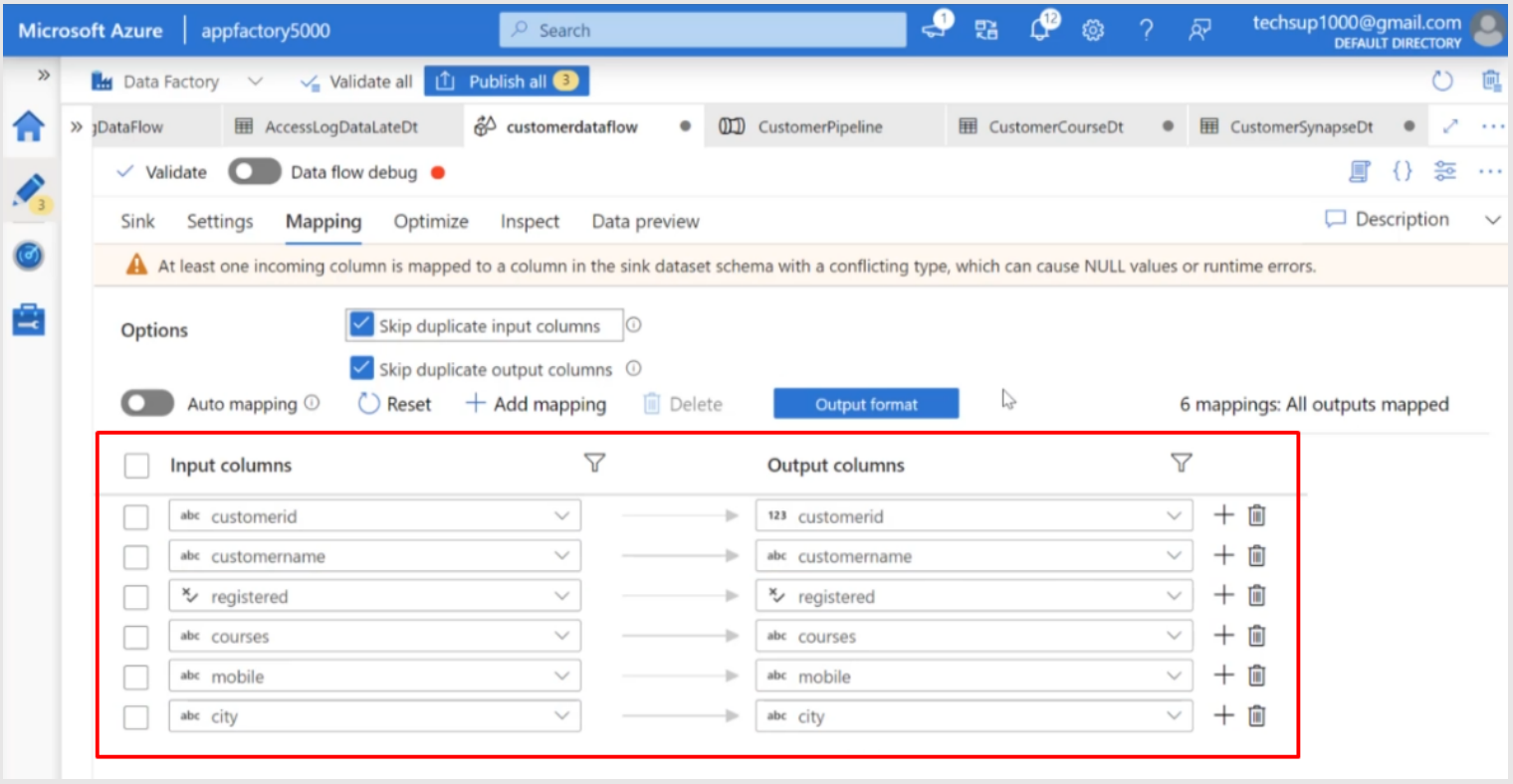

Microsoft Azure
appfactory5000
Search
techsup1000@gmail.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all 2
gpipeline
AccessLogDataFlow
AccessLogDataLateDt
customerdataflow
CustomerPipeline
CustomerCourseDt
Validate
Data flow debug
CustomerStream
FlattenStream
CustomerSynapse
Flatten settings
Optimize
Inspect
Data preview
Description
Options
Skip duplicate input columns
Skip duplicate output columns
Input columns *
Reset
Add mapping
Delete
6 mappings: All inputs mapped
CustomerStream's column
Name as
courses
customerid
customername
registered
details.mobile
details.city
mobile
city


Microsoft Azure
appfactory5000
Search
techsup1000@gmail.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all 2
Accesslogpipeline
AccessLogDataFlow
AccessLogDataLateDt
customerdataflow
Validate
Data flow debug
CustomerStream
Import data from CustomerCourseDt
FlattenStream
Unrolling arrays from courses to courses with columns 'customerid, customername, registered, courses'
sink1
Columns:
4 total
Sink
Settings
Mapping
Optimize
Inspect
Data preview
Description
Output stream name *
CustomerSynapse
Learn more
Incoming stream *
FlattenStream
Sink type *
Dataset
Inline
Cache
Dataset *
Select...
New
Options
Allow schema drift
Validate schema

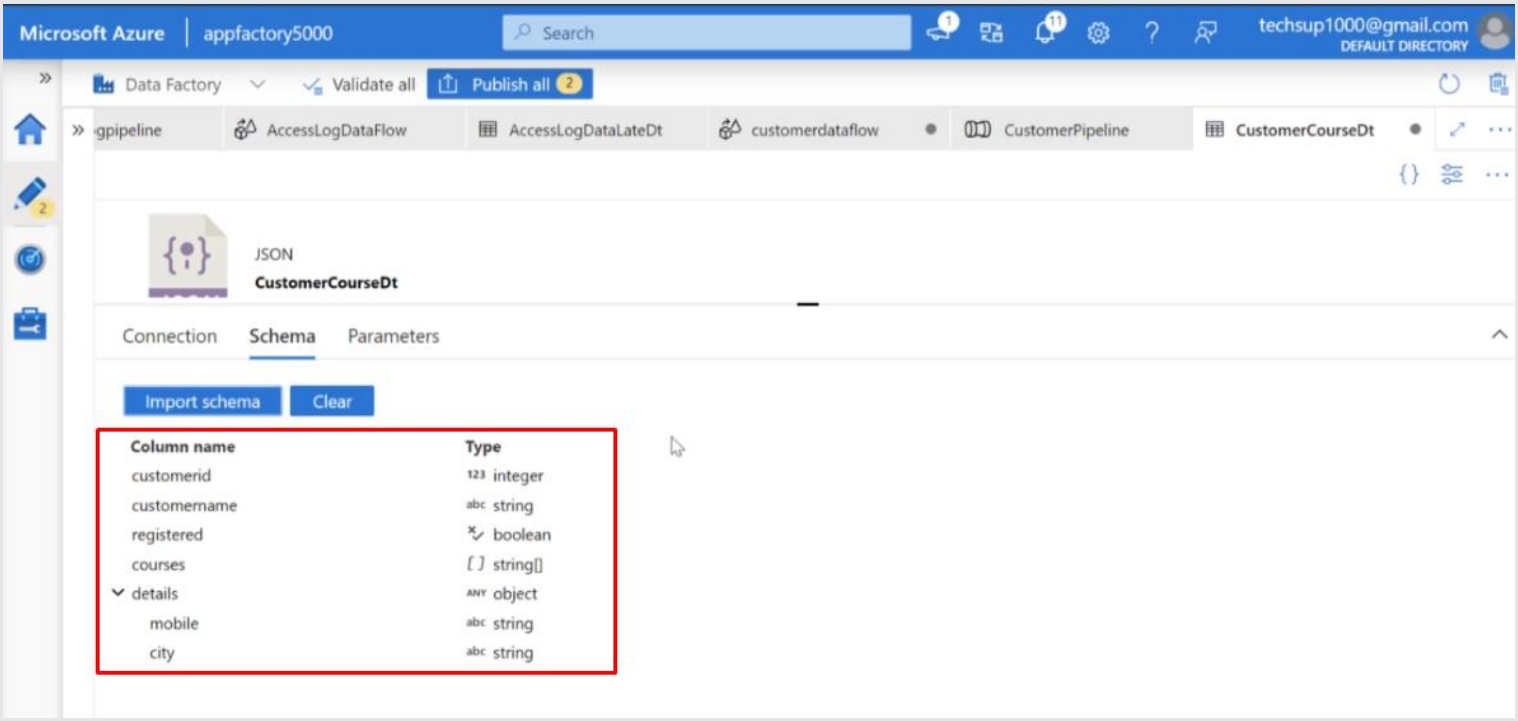
El dataset de destino apunta al Azure Synapse



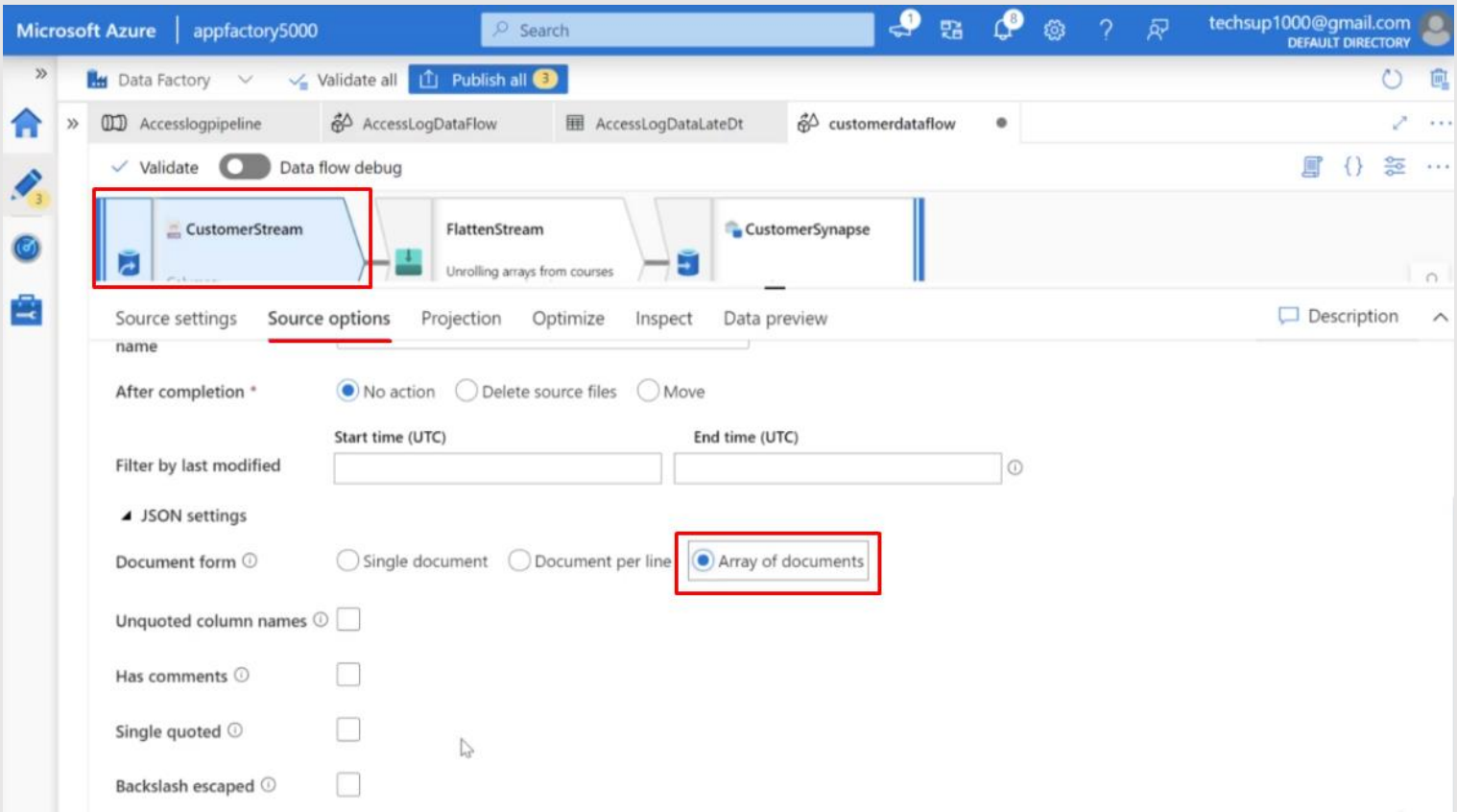
Revisamos el mapeo de las columnas en la caja de destino



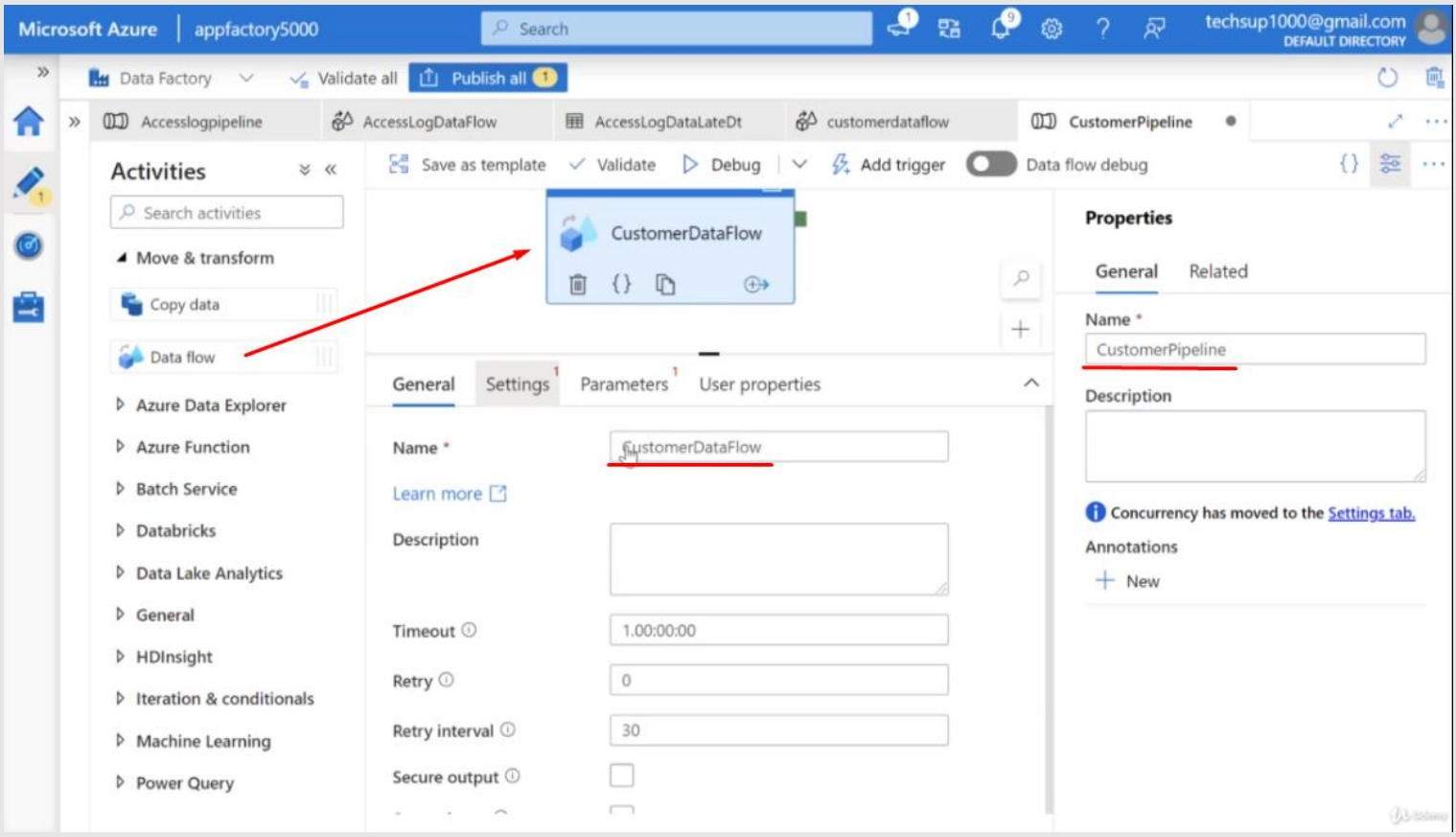
Si revisamos los tipos de datos que toma el archivo JSON Object desde la fuente

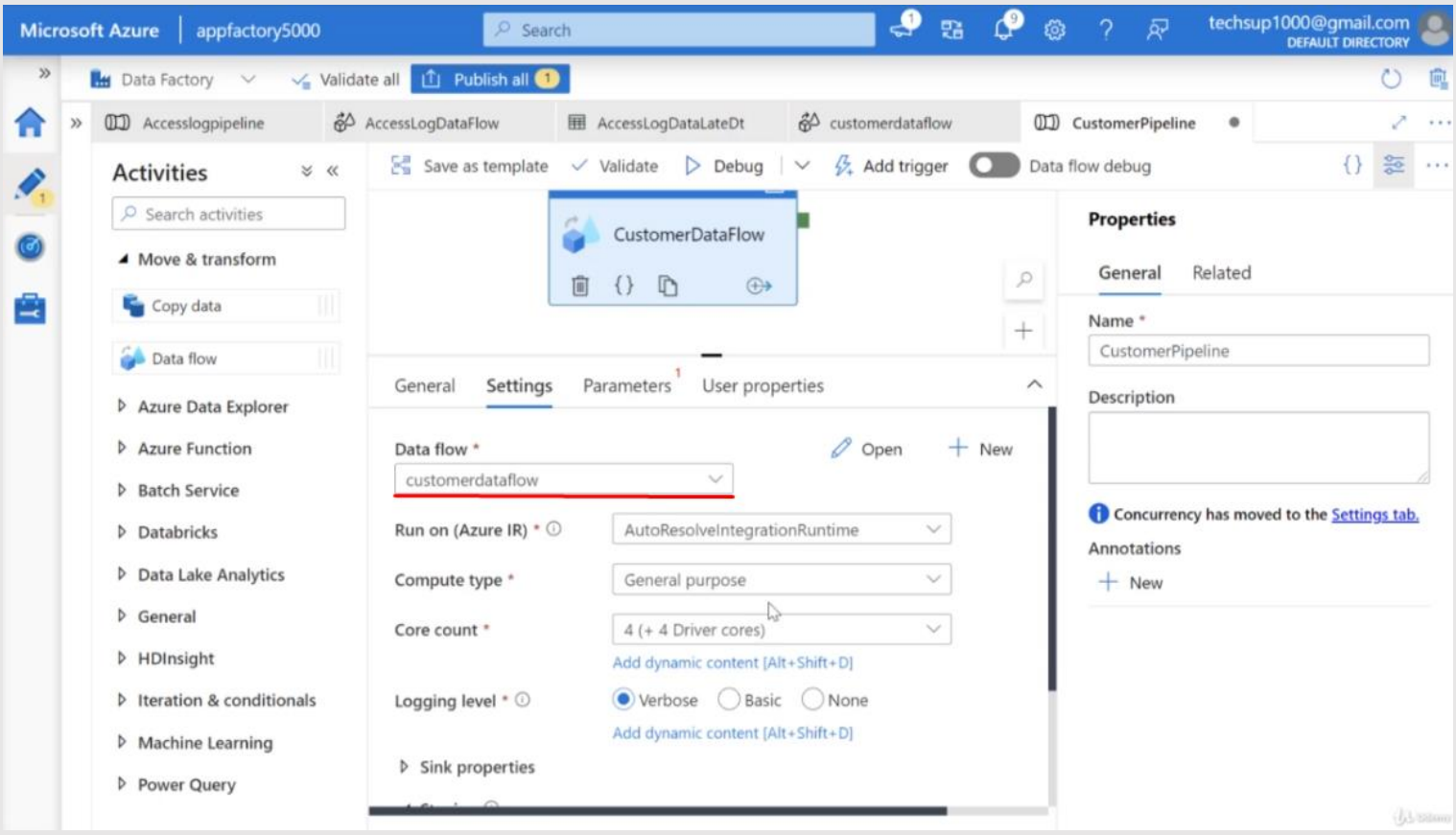


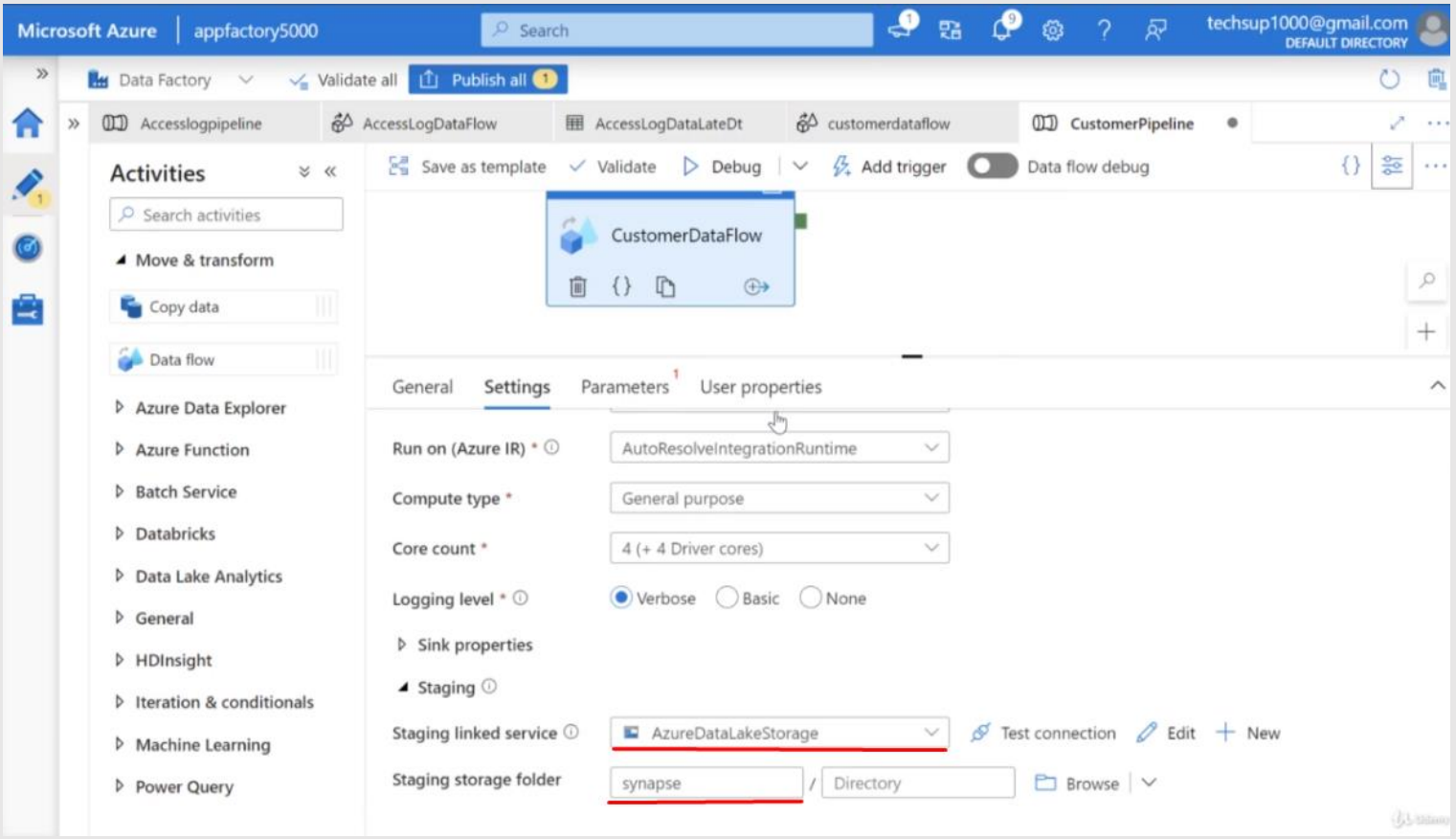
Y desde el paso fuente indicamos que corresponde a un archivo JSON Array



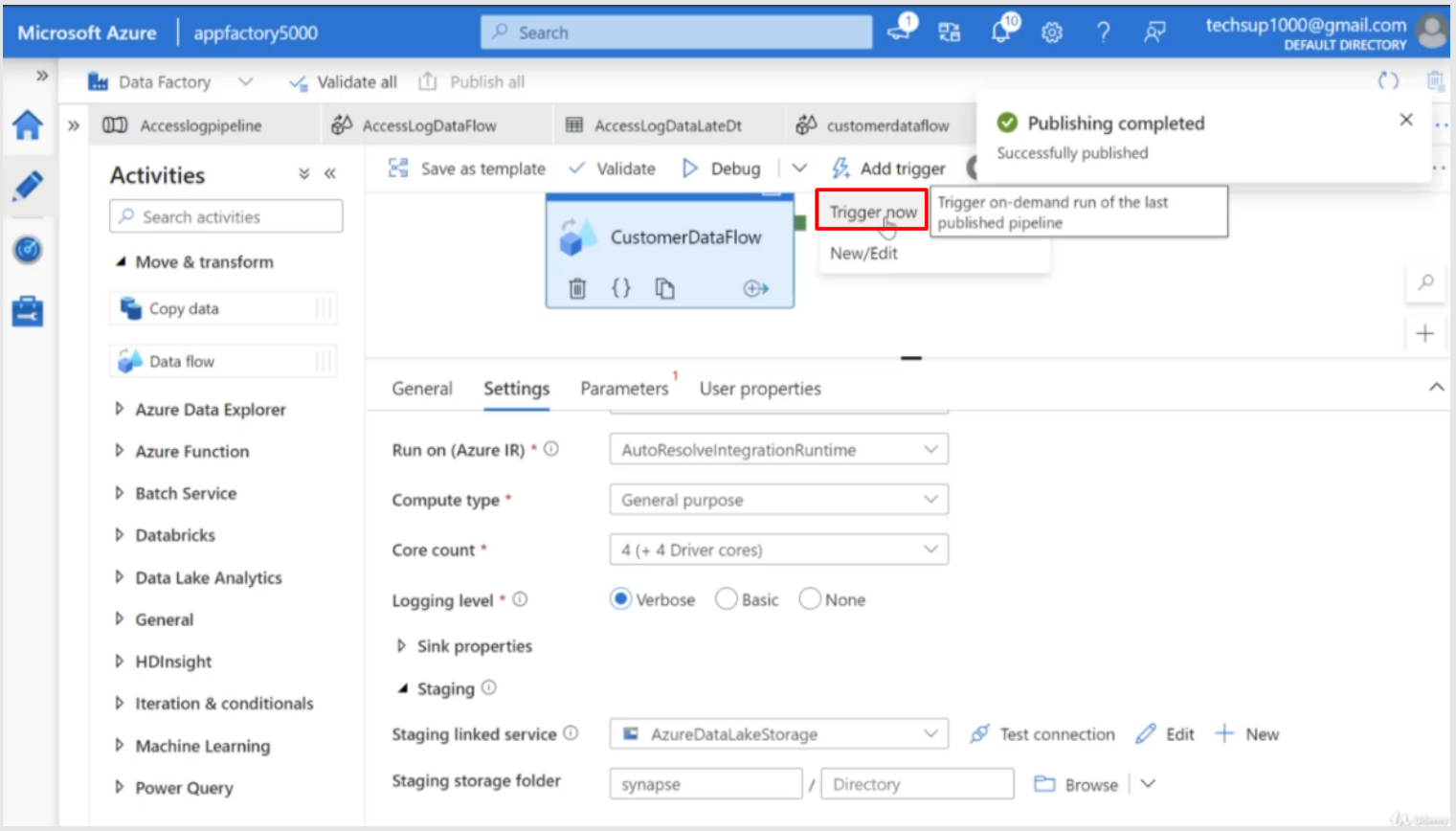
Para ejecutar nuestro Data Flow creamos un nuevo Pipeline







Validamos, **Publicamos** y **Ejecutamos**



Esta es la estructura de la tabla a la cual serán cargados los datos

```
drop table customercourse

CREATE TABLE [customercourse]
(
[customerid] int,
[customername] varchar(200),
[registered] BIT,
[courses] varchar(200),
[mobile] varchar(200),
[city] varchar(200)
)
```