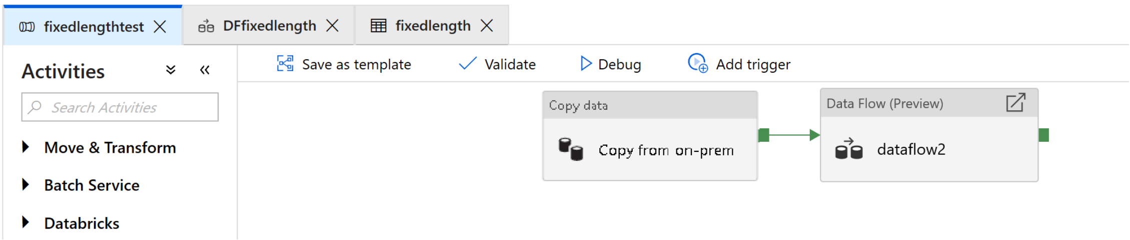
LAB - Process fixed-length text files by using Data Factory mapping data flows

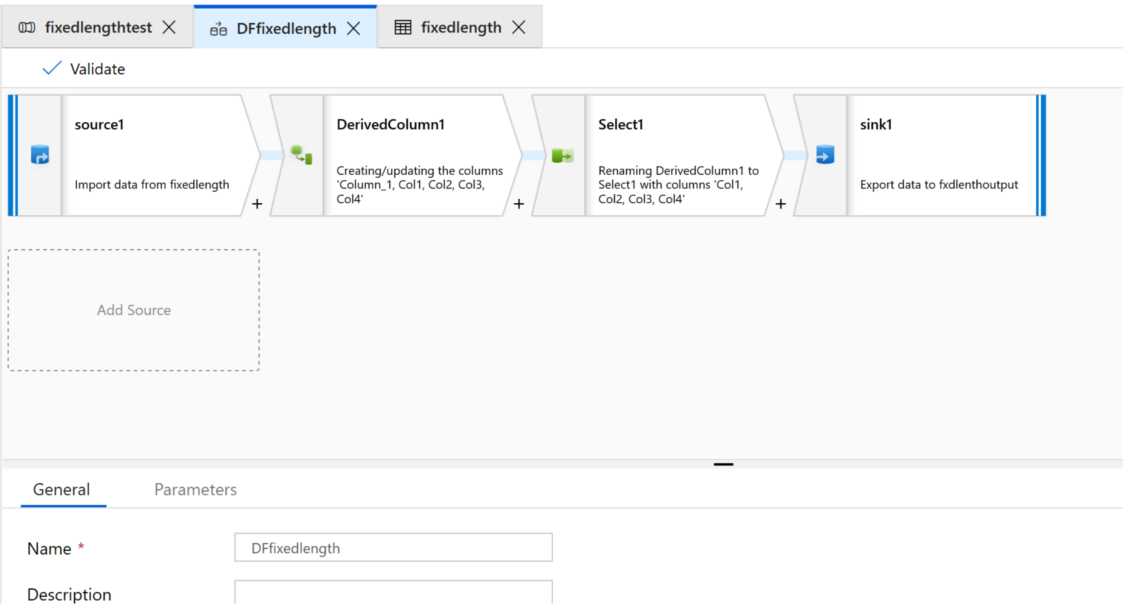
By using mapping data flows in Microsoft Azure Data Factory, you can transform data from fixed-width text files. In the following task, we'll define a dataset for a text file without a delimiter and then set up substring splits based on ordinal position.

Create a pipeline

1. Select **+New Pipeline** to create a new pipeline.
2. Add a data flow activity, which will be used for processing fixed-width files:



3. In the data flow activity, select **New mapping data flow**.
4. Add a Source, Derived Column, Select, and Sink transformation:


fixedlengthtest
DFfixedlength
fixedlength
Validate
source1
Import data from fixedlength
DerivedColumn1
Creating/updating the columns 'Column_1, Col1, Col2, Col3, Col4'
Select1
Renaming DerivedColumn1 to Select1 with columns 'Col1, Col2, Col3, Col4'
sink1
Export data to fxdlenthoutput
Add Source
General
Parameters

Name * DFfixedlength

Description

5. Configure the Source transformation to use a new dataset, which will be of the Delimited Text type.
6. Don't set any column delimiter or headers.

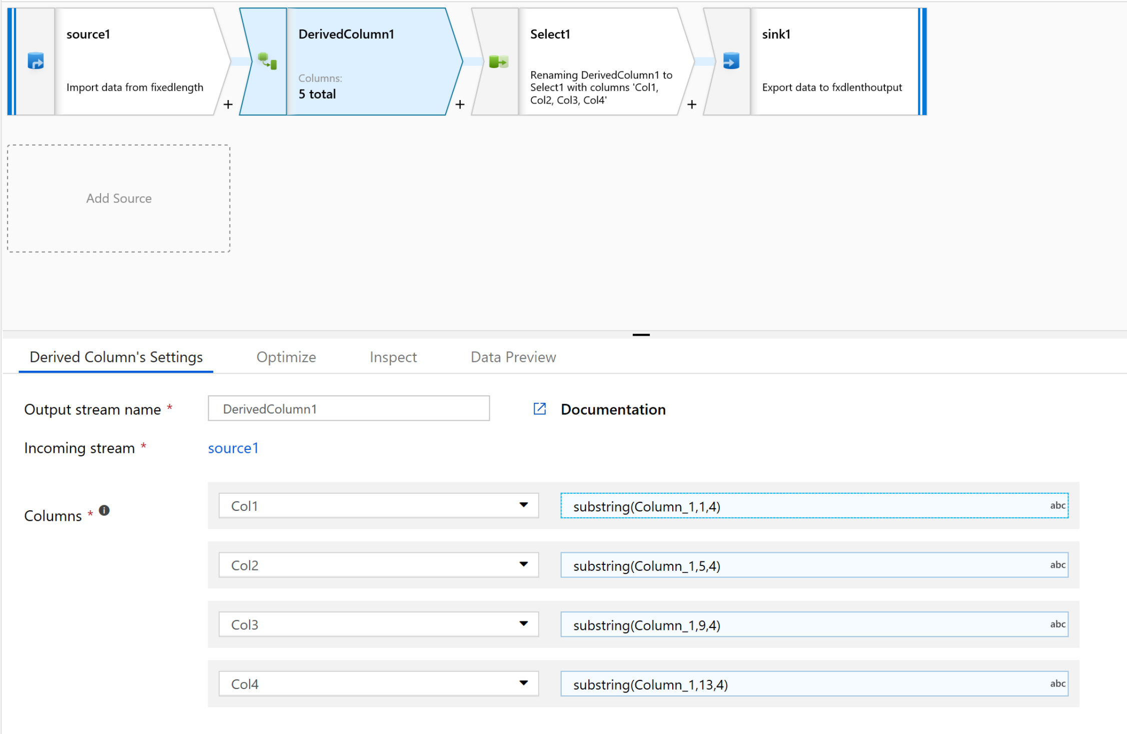
   Now we'll set field starting points and lengths for the contents of this file:

   Copy

```
1234567813572468
1234567813572468
1234567813572468
1234567813572468
1234567813572468
1234567813572468
1234567813572468
1234567813572468
1234567813572468
1234567813572468
1234567813572468
1234567813572468
1234567813572468
```

7. On the **Projection** tab of your Source transformation, you should see a string column that's named *Column_1*.
8. In the Derived column, create a new column.
9. We'll give the columns simple names like *col1*.
10. In the expression builder, type the following:

    `substring(Column_1,1,4)`


source1
Import data from fixedlength
DerivedColumn1
Columns:
5 total
Select1
Renaming DerivedColumn1 to Select1 with columns 'Col1, Col2, Col3, Col4'
sink1
Export data to fxdlenthoutput
Add Source
Derived Column's Settings
Optimize
Inspect
Data Preview
Documentation

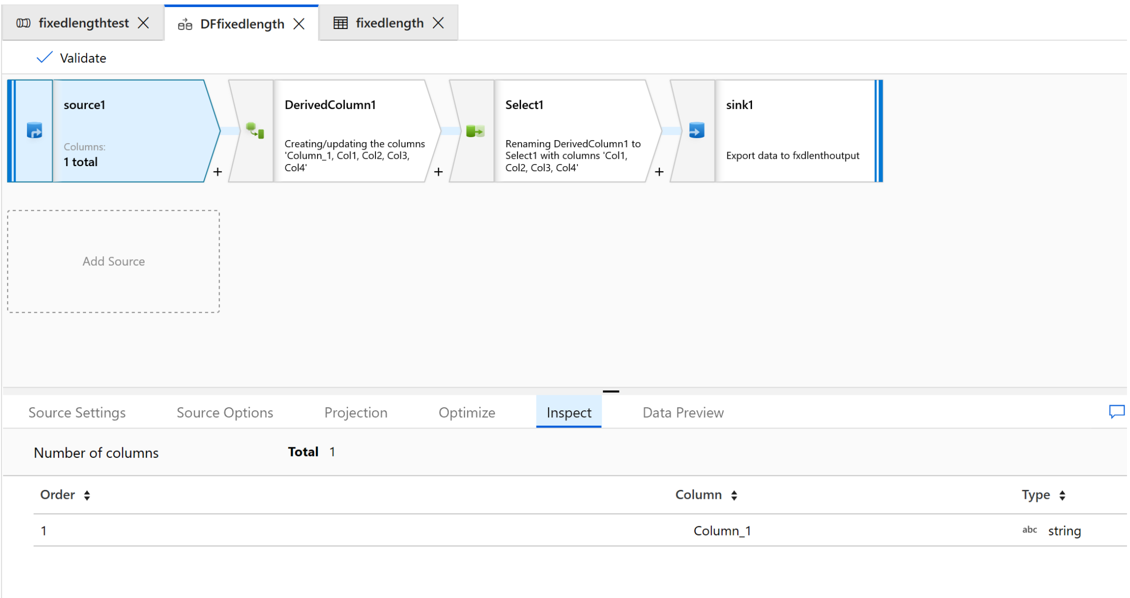
Output stream name * DerivedColumn1

Incoming stream * source1

Columns *

Col1	substring(Column_1,1,4)
Col2	substring(Column_1,5,4)
Col3	substring(Column_1,9,4)
Col4	substring(Column_1,13,4)

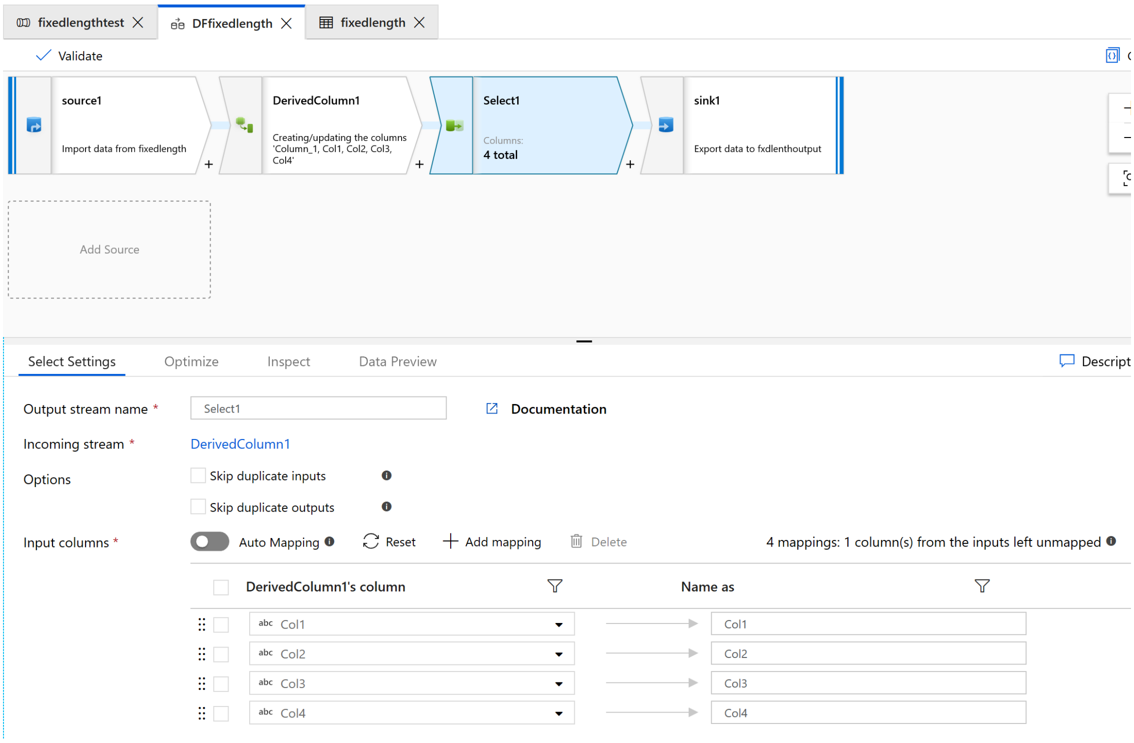
11. Repeat step 10 for all the columns you need to parse.
12. Select the **Inspect** tab to see the new columns that will be generated:



Number of columns **Total** 1

Order	Column	Type
1	Column_1	abc string

13. Use the Select transform to remove any of the columns that you don't need for transformation:



Output stream name * Select1

Incoming stream * DerivedColumn1

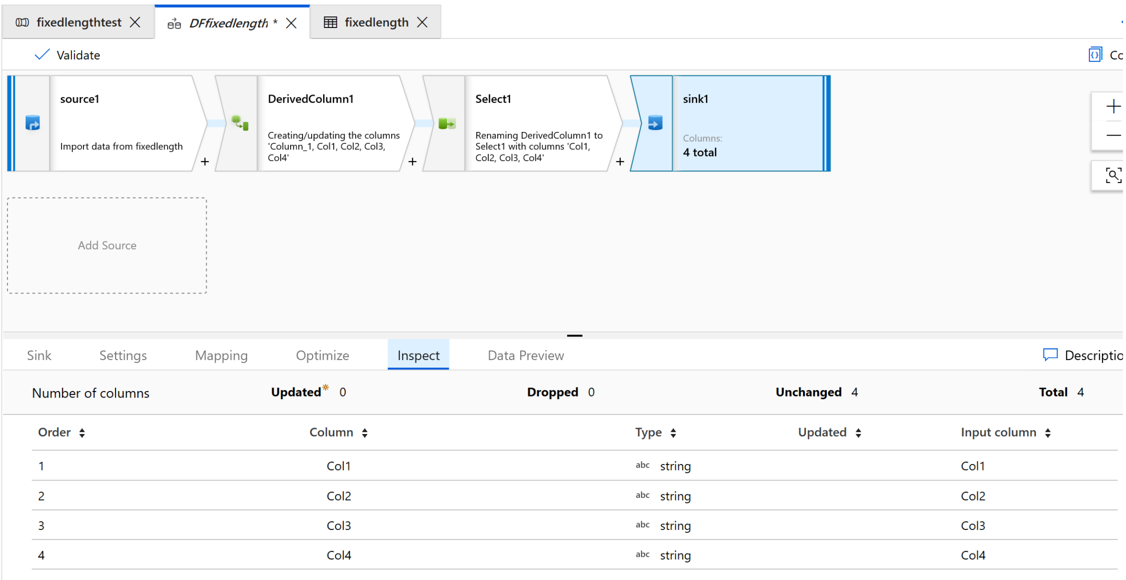
Options □ Skip duplicate inputs

□ Skip duplicate outputs

Input columns * Auto Mapping · Reset · Add mapping · Delete · 4 mappings: 1 column(s) from the inputs left unmapped

DerivedColumn1's column	Name as
abc Col1	Col1
abc Col2	Col2
abc Col3	Col3
abc Col4	Col4

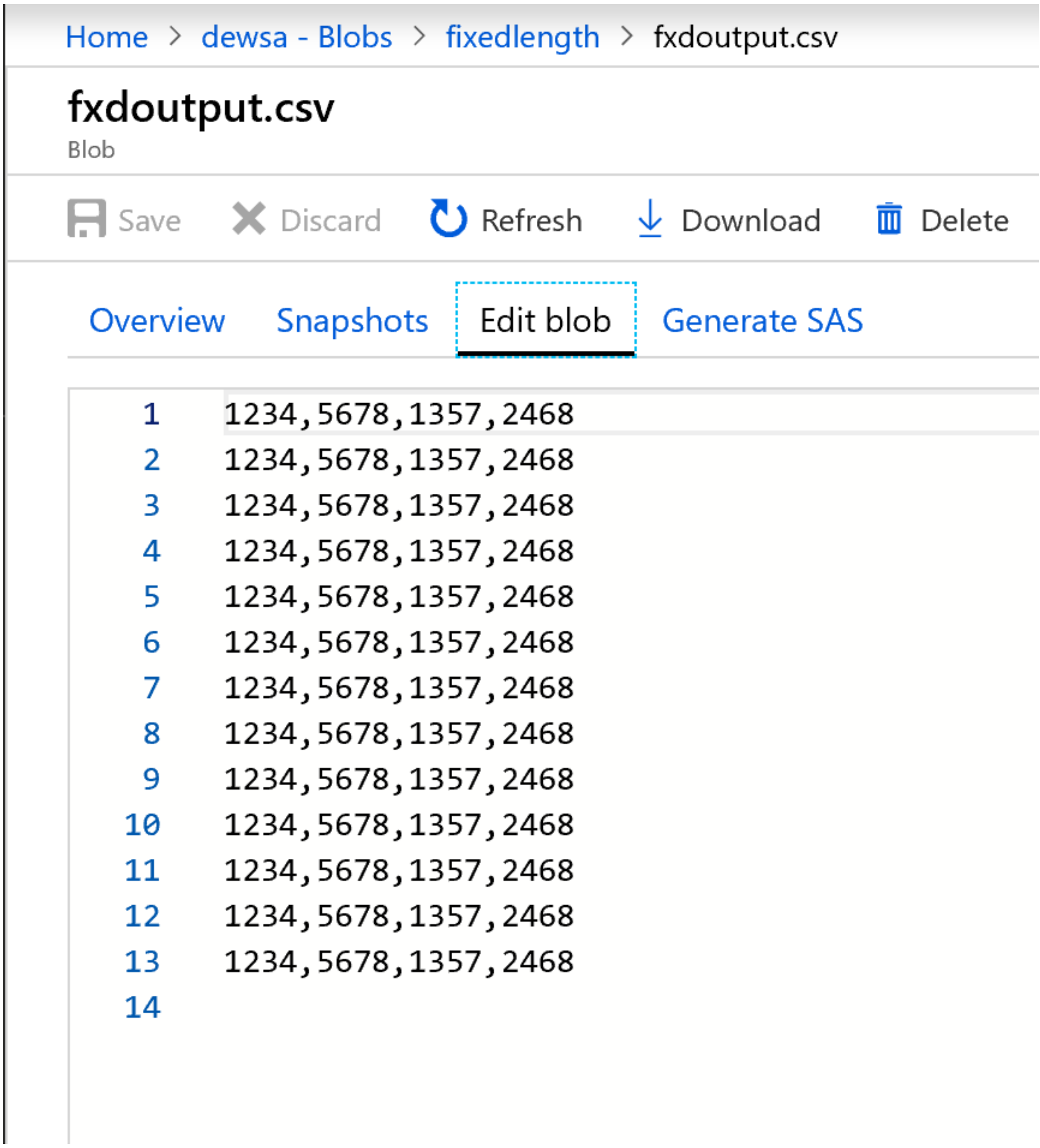
14. Use Sink to output the data to a folder:


fixedlengthtest
DFfixedlength *
fixedlength
Validate
source1
Import data from fixedlength
DerivedColumn1
Creating/updating the columns 'Column_1, Col1, Col2, Col3, Col4'
Select1
Renaming DerivedColumn1 to Select1 with columns 'Col1, Col2, Col3, Col4'
sink1
Columns:
4 total
Add Source
Sink
Settings
Mapping
Optimize
Inspect
Data Preview
Descriptio

Number of columns **Updated*** 0 **Dropped** 0 **Unchanged** 4 **Total** 4

Order	Column	Type	Updated	Input column
1	Col1	abc string		Col1
2	Col2	abc string		Col2
3	Col3	abc string		Col3
4	Col4	abc string		Col4

Here's what the output looks like:



```
1234,5678,1357,2468
1234,5678,1357,2468
1234,5678,1357,2468
1234,5678,1357,2468
1234,5678,1357,2468
1234,5678,1357,2468
1234,5678,1357,2468
1234,5678,1357,2468
1234,5678,1357,2468
1234,5678,1357,2468
1234,5678,1357,2468
1234,5678,1357,2468
1234,5678,1357,2468

```

The fixed-width data is now split, with four characters each and assigned to Col1, Col2, Col3, Col4, and so on. Based on the preceding example, the data is split into four columns.