Pipeline Template Import

There are two complete pipelines included as source code with this book. They can be imported directly into your existing Azure Data Factory as pipeline templates. Below are the steps you need to follow to import these templates into your factory.

# Importing pipeline templates

1. From the ADF browser UI, go to New > Pipeline > Import from pipeline template (figure 1).

Graphical user interface, application, Word

Description automatically generated

Figure . New pipeline from template

1. Select the template zip file from your file explorer (figure 2)



Figure . Select template zip file

1. After selecting the file, you will be prompted to choose Linked Services so that you can enter your own credentials and the file or database locations for the source data in each of the samples. You can either use an existing Linked Service or create new ones. The location of the sample data is included below in each sample.
2. When you are finished with the Linked Service selection, click “Use this template” and ADF will add the pipeline with the datasets and data flows to your factory (figure 3).

Graphical user interface, application

Description automatically generated

Figure . Final step in pipeline creation from template is to click "Use this template".

# Data deduplication

The data deduplication template is in file dedupePipeline.zip. There are two data flows included in this pipeline that represent the complete projects from the book where we created a distinct rows data flow and a deduplication scoring data flow. The source data for the distinct rows data flow uses the movie ranking database CSV file here: <https://github.com/kromerm/adfdataflowdocs/blob/master/sampledata/moviesDB.csv>. You can download that file and store it in your blob store container and then point to it using the source linked service prompt when creating the pipeline from the template.

The deduplication matching data flow uses sample names data from a random name generator [here](https://eforexcel.com/wp/downloads-16-sample-csv-files-data-sets-for-testing/). You can download a CSV from there and store it in your blob store container, then point to it with the linked service used for the deduplication matching data flow.

# Slowly changing dimensions

The slowly changing dimension sample is in the file called SCD Pipeline2.zip. The sample file used as the source for new products can be downloaded here and stored in your ADLS Gen2 storage container: <https://github.com/kromerm/adfdataflowdocs/blob/master/sampledata/Products.csv>. You’ll then point to that location using the linked service when creating the pipeline from the template.

To use the database sources, you will need create a linked service that points to an Azure SQL Database and load the products dimension table from the public Adventure Works sample SQL Server database: <https://docs.microsoft.com/en-us/sql/samples/adventureworks-install-configure>.