Context:

In this exercise, you'll use the Data Factory user interface (UI) to create a pipeline that copies and transforms data from an Azure Data Lake Storage Gen2 source to a Data Lake Storage Gen2 sink (both allowing access to only selected networks) by using mapping data flow in Data Factory Managed Virtual Network.

Learning Outcomes:

1) Create a data factory
2) Create a pipeline with a data flow activity.
3) Build a mapping data flow
4) Monitor a data flow activity.

Steps to be performed

1) Create an Azure Data Lake Gen2 storage account
2) Download the dataset in the link below and upload it to your Storage account
   https://raw.githubusercontent.com/djpmsft/adf-ready-demo/master/moviesDB.csv
3) 3) Create a Data Factory in any region.
4) Create an Azure Integration Runtime in the Data Factory and enable Virtual Network Configuration
5) Create a pipeline in the Data Factory with Mapping Data Flow
6) Configure the data flow to perform the following transformations
   a) Source transformation to read the dataset file from the Gen2 storage account
   b) A filter transformation to retrieve all movies from the years 1910-2000 under the genre "Comedy"
   c) An aggregate transformation to return the "Rating" of all the records retrieved in the previous step
   d) A sink transformation to write the output of the previous transformation into a folder in the same storage account.

Note:

● Other required values can be set at your discretion.
● Submission of this assessment shall be done in the form of a pdf document containing the labeled screenshots as outlined in the marks distribution section.