**Case Study 2 Solution**

1. Create Azure Blob Storage
2. Click Storage Accounts in Azure Portal
3. Click on Create, Select Subscription and Resource Group, Specify Storage Account Name, Select Region and Click on Create button.
4. Create Container under Storage Account and upload 2016\_medals.csv
5. Create Azure Data Lake
   1. Click Storage Accounts in Azure Portal.
   2. Click on Create, Select Subscription and Resource Group, Specify Storage Account Name and Select Region.
   3. Click on Advanced tab and Under Data Lake Storage Gen2 enable hierarchical namespace option.
   4. Click on Create button.
6. Create Data factory and Pipeline
   1. Create Data Factory.
   2. Create New Mapping Data Flow.
   3. Click Add Source, Under Source settings create dataset for 2016\_medals.csv file present in Azure Blob storage.
   4. Add Conditional Split Transformation in to the Canvas.
   5. Specify name,Split condition as follows for each medal type :
      1. Provide Stream names as Gold,Silver and Bronze with conditions using equals(<Select column by selecting input schema (Medal)>, ‘<Specify condition in single code like Gold,Silver or Bronze>’)
      2. Create 3 split conditions as we have Gold,Silver and Bronze values.
   6. Create sink Datasets by enabling output to single file option for each of the split conditions we created.
   7. Publish the Data Flow.
7. Create a Pipeline, use Dataflow activity and select Dataflow created in previous step.
8. Execute the Pipeline.