**DATA ENGINEERING**

**PipeLines Built on day 14th**

**1.**



The above picture is of a dataflow activity which is used to create a stored event trigger

Inside the dataflow activity we have the sort function which is used to display the details of a csv file which is of employee sorted based on the salary in descending order.

The trigger is activated when ever there are any changes in the input file or source we are using

Note:we are using the parameters to provide the folder and the file name not by hardcoding the names instead we use the triggerbody in the pipeline trigger settings

So whenever there is a change in the source file the trigger is activated and performs the calculation on the present data and stores it in the Destination

2.



For the second pipeline built we created a Azure Sql database Dataset in which we will store the sql data we retrieve from the sql server and the database we created using the sample adventureworks database .We used the copydata activity to achieve the requirement

We define a query from which we will get the data from the Azure SQL Database linked service created when the dataset of Azure Sql database is created and then the result is stored at the destination

3,4.



For the third and fourth pipeline we use data from an api called rebrickable which provides us with some amount of csv files.For the 3rd pipeline we use the https linked service from which we will be creating dataset which will hold the data files we get from the api.So now we have options called baseurl which will have the api address and the relative path will contain the file name which will be retrieved and stored at the destination once debuged

For the fourth Now suppose we need to retrieve the multiple files we can create a variable which is array and store all the names of the files .Using the foreach loop activity we can now get the files which are compressed in .gz format so we will choose compression type also as .gz and the type as faster now with the help of parameter and variable we will retrieve the multiples files from the api .Basically copy data is placed in the foreach to retrieve the files specified in the variable

5.



The 5th pipeline is kind of extension of 3rd and 4th now the array variable harcoding is not a feasible solution in a realtime world application .Suppose we have a sql table which gets updated and those files should be downloaded or we have a required that all the files in the table of a database should be downloaded then we can make use of the Azure sql database we created earlier to do lookup by using a query as "select filename from the info" table

Note:

The table should be created or be present

Now the result of the lookup is loaded into the foreach activity from which the "@activity('Lookup1').output.value" and then further processed in the copy activity which uses the httpserver dataset created earlier with the parameter but the variable is replaced by the output of the foreach "@item().filename" which are names of the files we will be retrieving and storing