Data Lakes and Data Integration
Project – part 1

The objective of this project is to build a full pipeline processing a dataset (link : https://drive.google.com/file/d/1F63LhH7LycfQPiFpjVTqg5mWe2t_YiLy/view?usp=sharing) composed of stocks data following these steps :

1. Import data of the stocks that is available in this link into an azure storage solution using the script that you have developed in lab 1, optional, you can set an activity that takes
2. Transform the csv files into one csv file and add the stock name column to it, and then store it into a data lake with an Azure batch activity
3. Create a copy activity that takes the data from the csv file and put it into a SQL database
4. Create a databricks activity that connect to this sql database and code the following functions :
    a) Daily return rate : a function that get a stock name, a start and end date and output the daily return date of this stock during this period
    b) Moving average : a function that takes a stock name, a start and end data, and a number of moving points (5 points for example) and return a new dataframe with the applied moving average over the opening price column
    c) The databricks activity must output the results of the notebook into a storage account

- Once you developed your full pipeline, record it (using loom or any other screen recording solution) by showing :
    - A slide representing your architecture
    - subscription details and email
    - The execution details

- Also, set a github repository where you will put the code that you have used in the python batch and databricks activities