**Assignment:**

Create an ETL pipeline that gets the data from a CSV file in google cloud storage and save that data to BigQuery table. If there are any timestamp conversions to UTC needed please add the transformations.

Use the following New york citibank bike data set from the following location. Use probably a couple of files for your testing.

<https://s3.amazonaws.com/tripdata/index.html>

You may create the table in BigQuery with the following schema elements (ref : <https://www.citibikenyc.com/system-data>)

* Trip Duration (seconds)
* Start Time and Date
* Stop Time and Date
* Start Station Name
* End Station Name
* Station ID
* Station Lat/Long
* Bike ID
* User Type (Customer = 24-hour pass or 3-day pass user; Subscriber = Annual Member)
* Gender (Zero=unknown; 1=male; 2=female)
* Year of Birth

You have a wide choice of accomplishing this exercise. You can use any one of following ETL technologies

1. CDAP or Data Fusion
2. Airflow or Composer
3. DataFlow or DataProc or Running Spark on local machine

Time: 1 week

Expectations:

Give a 30 minute demo at the end of 1 week.