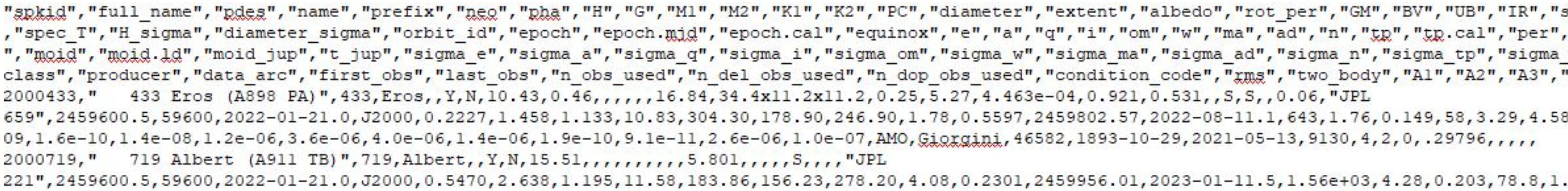
Near Earth Objects

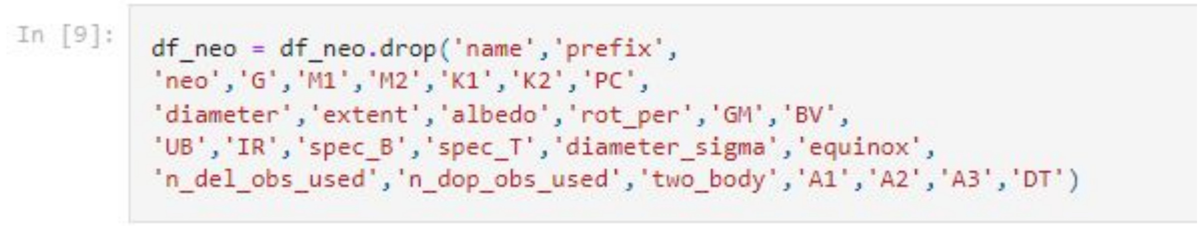
ETL Process

Download CSV Data



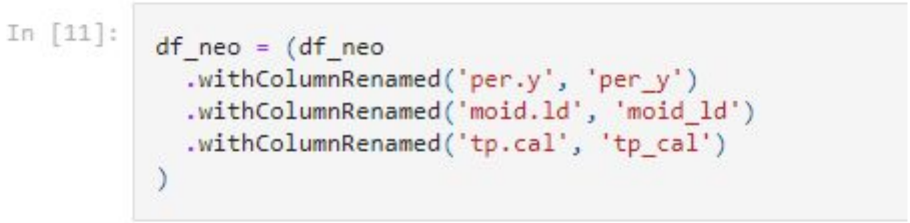
```
"spkid","full_name","pdes","name","prefix","neo","pha","H","G","M1","M2","K1","K2","PC","diameter","extent","albedo","rot_per","GM","BV","UB","IR","s
,"spec_T","H_sigma","diameter_sigma","orbit_id","epoch","epoch.mjd","epoch.cal","equinox","e","a","q","i","om","w","ma","ad","n","tp","tp.cal","per",
","moid","moid.ld","moid_jup","t_jup","sigma_e","sigma_a","sigma_q","sigma_i","sigma_om","sigma_w","sigma_ma","sigma_ad","sigma_n","sigma_tp","sigma_
class","producer","data_arc","first_obs","last_obs","n_obs_used","n_del_obs_used","n_dop_obs_used","condition_code","rms","two_body","A1","A2","A3","
2000433,"   433 Eros (A898 PA)",433,Eros,,Y,N,10.43,0.46,,,,,,16.84,34.4x11.2x11.2,0.25,5.27,4.463e-04,0.921,0.531,,S,S,,0.06,"JPL
659",2459600.5,59600,2022-01-21.0,J2000,0.2227,1.458,1.133,10.83,304.30,178.90,246.90,1.78,0.5597,2459802.57,2022-08-11.1,643,1.76,0.149,58,3.29,4.58
09,1.6e-10,1.4e-08,1.2e-06,3.6e-06,4.0e-06,1.4e-06,1.9e-10,9.1e-11,2.6e-06,1.0e-07,AMO,Giorgini,46582,1893-10-29,2021-05-13,9130,4,2,0,.29796,,,,,
2000719,"   719 Albert (A911 TB)",719,Albert,,Y,N,15.51,,,,,,,,,,5.801,,,,,S,,,,"JPL
221",2459600.5,59600,2022-01-21.0,J2000,0.5470,2.638,1.195,11.58,183.86,156.23,278.20,4.08,0.2301,2459956.01,2023-01-11.5,1.56e+03,4.28,0.203,78.8,1.
```

Transformation



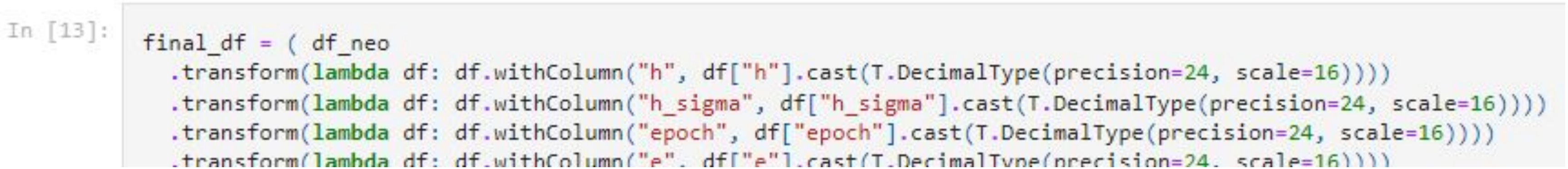
In [9]:

```
df_neo = df_neo.drop('name','prefix',
'neo','G','M1','M2','K1','K2','PC',
'diameter','extent','albedo','rot_per','GM','BV',
'UB','IR','spec_B','spec_T','diameter_sigma','equinox',
'n_del_obs_used','n_dop_obs_used','two_body','A1','A2','A3','DT')
```



In [11]:

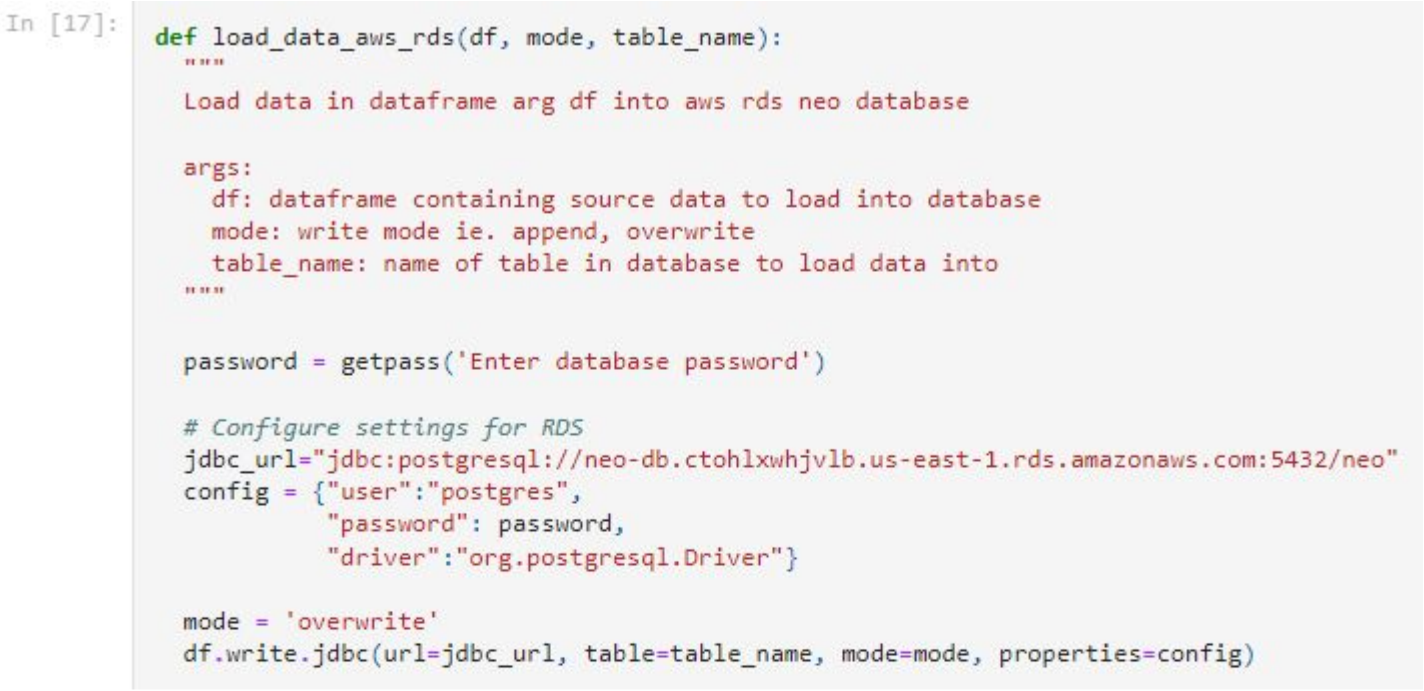
```
df_neo = (df_neo
  .withColumnRenamed('per.y', 'per_y')
  .withColumnRenamed('moid.ld', 'moid_ld')
  .withColumnRenamed('tp.cal', 'tp_cal')
)
```



In [13]:

```
final_df = ( df_neo
  .transform(lambda df: df.withColumn("h", df["h"].cast(T.DecimalType(precision=24, scale=16))))
  .transform(lambda df: df.withColumn("h_sigma", df["h_sigma"].cast(T.DecimalType(precision=24, scale=16))))
  .transform(lambda df: df.withColumn("epoch", df["epoch"].cast(T.DecimalType(precision=24, scale=16))))
  .transform(lambda df: df.withColumn("e", df["e"].cast(T.DecimalType(precision=24, scale=16))))
```

Load



In [17]:

```
def load_data_aws_rds(df, mode, table_name):
  """
  Load data in dataframe arg df into aws rds neo database

  args:
    df: dataframe containing source data to load into database
    mode: write mode ie. append, overwrite
    table_name: name of table in database to load data into
  """

  password = getpass('Enter database password')

  # Configure settings for RDS
  jdbc_url="jdbc:postgresql://neo-db.ctohlxwhjvlb.us-east-1.rds.amazonaws.com:5432/neo"
  config = {"user":"postgres",
            "password": password,
            "driver":"org.postgresql.Driver"}

  mode = 'overwrite'
  df.write.jdbc(url=jdbc_url, table=table_name, mode=mode, properties=config)
```