

```
1 ###
2 from pyspark.sql import SparkSession
3 from pyspark.sql.functions import col, lit
4 ###
5 appName = "IndianStockAnalysis"
6 # Create a Spark session
7 spark = SparkSession.builder.appName(appName).
   getOrCreate()
8 ###
9 # Load CSV files into PySpark DataFrames
10 adaniports_df = spark.read.csv("hdfs://localhost:9000/
   /user/input/IndianStocks/ADANIPORTS.csv", header=True
   )
11 asianpaint_df = spark.read.csv("hdfs://localhost:9000/
   /user/input/IndianStocks/ASIANPAINT.csv", header=True
   )
12 axisbank_df = spark.read.csv("hdfs://localhost:9000/
   user/input/IndianStocks/AXISBANK.csv", header=True)
13 bajaj_auto_df = spark.read.csv("hdfs://localhost:9000/
   /user/input/IndianStocks/BAJAJ-AUTO.csv", header=True
   )
14 bajajfinsv_df = spark.read.csv("hdfs://localhost:9000/
   /user/input/IndianStocks/BAJAJFINSV.csv", header=True
   )
15 bajfinance_df = spark.read.csv("hdfs://localhost:9000/
   /user/input/IndianStocks/BAJFINANCE.csv", header=True
   )
16 bhartiartl_df = spark.read.csv("hdfs://localhost:9000/
   /user/input/IndianStocks/BHARTIARTL.csv", header=True
   )
17 bpcl_df = spark.read.csv("hdfs://localhost:9000/user/
   input/IndianStocks/BPCL.csv", header=True)
18 britannia_df = spark.read.csv("hdfs://localhost:9000/
   user/input/IndianStocks/BRITANNIA.csv", header=True)
19 cipla = spark.read.csv("hdfs://localhost:9000/user/
   input/IndianStocks/CIPLA.csv", header=True)
20 coalindia = spark.read.csv("hdfs://localhost:9000/
   user/input/IndianStocks/COALINDIA.csv", header=True)
21 drreddy = spark.read.csv("hdfs://localhost:9000/user/
   input/IndianStocks/DRREDDY.csv", header=True)
22 eichermot = spark.read.csv("hdfs://localhost:9000/
```

```
22 user=input/IndianStocks/EICHERMOT.csv", header=True)
23 gail = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/GAIL.csv", header=True)
24 grasim = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/GRASIM.csv", header=True)
25 hcltech = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/HCLTECH.csv", header=True)
26 hdfc = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/HDFC.csv", header=True)
27 hdfcbank = spark.read.csv("hdfs://localhost:9000/user
    /input/IndianStocks/HDFCBANK.csv", header=True)
28 heromotoco = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/HEROMOTOCO.csv", header=True)
29 hindalco = spark.read.csv("hdfs://localhost:9000/user
    /input/IndianStocks/HINDALCO.csv", header=True)
30 hindunilvr = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/HINDUNILVR.csv", header=True)
31 icicibank = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/ICICIBANK.csv", header=True)
32 indusindbk = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/INDUSINDBK.csv", header=True)
33 infratel = spark.read.csv("hdfs://localhost:9000/user
    /input/IndianStocks/INFRATEL.csv", header=True)
34 infy = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/INFY.csv", header=True)
35 ioc = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/IOC.csv", header=True)
36 itc = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/ITC.csv", header=True)
37 jswsteel = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/JSWSTEEL.csv", header=True)
38 kotakbank = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/KOTAKBANK.csv", header=True)
39 lt = spark.read.csv("hdfs://localhost:9000/user/input
    /IndianStocks/LT.csv", header=True)
40 maruti = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/MARUTI.csv", header=True)
41 mm = spark.read.csv("hdfs://localhost:9000/user/input
    /IndianStocks/MM.csv", header=True)
42 nestleind = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/NESTLEIND.csv", header=True)
```

```

43 ntpc = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/NTPC.csv", header=True)
44 ongc = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/ONGC.csv", header=True)
45 powergrid = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/POWERGRID.csv", header=True)
46 reliance = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/RELIANCE.csv", header=True)
47 sbi = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/SBI.csv", header=True)
48 shreecem = spark.read.csv("hdfs://localhost:9000/user
    /input/IndianStocks/SHREECEM.csv", header=True)
49 sunpharma = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/SUNPHARMA.csv", header=True)
50 tatamotors = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/TATAMOTORS.csv", header=True)
51 tatasteel = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/TATASTEEL.csv", header=True)
52 tcs = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/TCS.csv", header=True)
53 techm = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/TECHM.csv", header=True)
54 titan = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/TITAN.csv", header=True)
55 ultracemco = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/ULTRACEMCO.csv", header=True)
56 upl = spark.read.csv("hdfs://localhost:9000/user/input
    /IndianStocks/UPL.csv", header=True)
57 vedl = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/VEDL.csv", header=True)
58 wipro = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/WIPRO.csv", header=True)
59 zeel = spark.read.csv("hdfs://localhost:9000/user/
    input/IndianStocks/ZEEL.csv", header=True)
60 nifty50_all = spark.read.csv("hdfs://localhost:9000/
    user/input/IndianStocks/NIFTY50_all.csv", header=True
    )
61
62 ##
63 #Merging the dataframes
64 all_stocks = adaniports_df.union(asianpaint_df).union

```

```

64 (axisbank_df).union(bajaj_auto_df).union(
    bajajfinsv_df).union(bajfinance_df).union(
    bhartiartl_df).union(bpcl_df).union(britannia_df).
    union(cipla).union(coalindia).union(drreddy).union(
    eichermt).union(gail).union(grasim).union(hcltech).
    union(hdfc).union(hdfcbank).union(heromotoco).union(
    hindalco).union(hindunilvr).union(icicibank).union(
    indusindbk).union(infratel).union(infy).union(ioc).
    union(itc).union(jswsteel).union(kotakbank).union(lt
    ).union(maruti).union(mm).union(nestleind).union(
    ntpc).union(ongc).union(powergrid).union(reliance).
    union(sbin).union(shreecem).union(sunpharma).union(
    tatamotors).union(tatasteel).union(tcs).union(techm
    ).union(titan).union(ultracemco).union(upl).union(
    vedl).union(wipro).union(zeel).union(nifty50_all)
65 #%%
66 all_stocks.count()
67 #%%
68 #We do some transformation in some of the fields,
and get the fields we are interested.
69 all_stocks = all_stocks.select("Date", "Symbol", "
    Open", "Close", "Low", "High")
70 #%%
71 all_stocks = all_stocks.withColumn("Country", lit("
    India"))
72 #%%
73 all_stocks.where(all_stocks["Symbol"] == "LT").show(
    5)
74 #%%
75 #In case of failure, save the dataframe in single
file
76 all_stocks.coalesce(1).write.csv("/home/azureuser/
    bigdata/all_stocks/single_csv_file", header=True,
    mode='overwrite')
77 #%%
78 #Installing libraries to connect to azureSQL and
transfer data
79 !pip install sqlalchemy
80 !pip install pyodbc
81 !pip install pandas
82 #%%

```

```
83 #Connecting and testing connection to database
84 server = 'alinizarserver1.database.windows.net'
85 database = 'myfirstdatabase'
86 username = 'azureuser@alinizarserver1'
87 password = 'Ilove85workWonder69'
88 driver = 'ODBC Driver 17 for SQL Server'
89 table_name = 'STOCKS'
90 ###
91 from sqlalchemy import create_engine
92 connection_string = f'mssql+pyodbc://{username}:{password}@{server}:1433/{database}?driver={driver}'
93 # Create an engine and connect
94 engine = create_engine(connection_string)
95 connection = engine.connect()
96 ###
97 #Test query
98 result = connection.execute("SELECT @@VERSION")
99 for row in result:
100     print(row[0])
101 ###
102 import pandas as pd
103 #Insert dataframe into AzureSQL table
104 pd_dataframe = all_stocks.toPandas()
105
106 pd_dataframe.to_sql(table_name, engine, if_exists='
    replace', index=False)
107 print("Data transfer completed successfully.")
108
109 # Stop the Spark session
110 spark.stop()
```