Harshil Patel (0501)

Class Activity:

```
#Practical Exercise: Display the information where location is "New York" and revenue is less than 5.0
practice_result=hiveCtx.sql("select * from jsontable where location = 'New York' and revenue < 5.0")
practice_result.show()
```

```
+---------------+--------+---------------+-------+--------------------+
|           name|location|        address|revenue|           employees|
+---------------+--------+---------------+-------+--------------------+
|Branch Office 1|New York|123 Main Street|    3.2|[{name -> John Do...|
|Branch Office 3|New York| 789 Elm Street|    4.2|[{name -> David L...|
+---------------+--------+---------------+-------+--------------------+
```

```
!pip install pyspark
```

```
Collecting pyspark
  Downloading pyspark-3.5.1.tar.gz (317.0 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 317.0/317.0 MB 1.3 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: py4j==0.10.9.7 in /usr/local/lib/python3.10/dist-packages (from pyspark) (0.10.9.7)
Building wheels for collected packages: pyspark
  Building wheel for pyspark (setup.py) ... done
  Created wheel for pyspark: filename=pyspark-3.5.1-py2.py3-none-any.whl size=317488493 sha256=f0ddd2ecdf343059eb4b3b55fe31e1484d3bcc1b1
  Stored in directory: /root/.cache/pip/wheels/80/1d/60/2c256ed38dddce2fdd93be545214a63e02fbd8d74fb0b7f3a6
Successfully built pyspark
Installing collected packages: pyspark
Successfully installed pyspark-3.5.1
```

```
from pyspark.sql import HiveContext, Row
```

```
from pyspark.sql import SparkSession
if __name__ == "__main__":
    #session=SparkSession.builder.appName("ex1").master("local[2]").getOrCreate()
    #dataFrameReader=session.read
    session=SparkSession.builder.enableHiveSupport().getOrCreate()
    sc=session.sparkContext
    hiveCtx = HiveContext(sc)
```

```
/usr/local/lib/python3.10/dist-packages/pyspark/sql/context.py:733: FutureWarning: HiveContext is deprecated in Spark 2.0.0. Please use
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/pyspark/sql/context.py:113: FutureWarning: Deprecated in 3.0.0. Use SparkSession.builder.getOrCr
  warnings.warn(
```

```
#Creating table from json file
import pandas
data=pandas.read_json("BranchOffice.json")
df=hiveCtx.createDataFrame(data)
df.registerTempTable("jsontable")
```

```
/usr/local/lib/python3.10/dist-packages/pyspark/sql/dataframe.py:329: FutureWarning: Deprecated in 2.0, use createOrReplaceTempView inst
  warnings.warn("Deprecated in 2.0, use createOrReplaceTempView instead.", FutureWarning)
```

```
result=hiveCtx.sql("SELECT * FROM jsontable")
result.show(5,truncate=False)
```

```
-----------------------------------------------------------------------------------------------------------------------+
                                                                                                                       |
-----------------------------------------------------------------------------------------------------------------------+
 -> Jane Smith, position -> Assistant Manager, phone -> 555-987-6543, email -> jane.smith@example.com}]                |
le.com}, {name -> Emily Brown, position -> Assistant Manager, phone -> 555-876-5432, email -> emily.brown@example.com}] |
me -> Sarah Taylor, position -> Assistant Manager, phone -> 555-654-3210, email -> sarah.taylor@example.com}]          |
inez@example.com}, {name -> Jessica Nguyen, position -> Assistant Manager, phone -> 555-543-2109, email -> jessica.nguyen@example.com}]|
om}, {name -> Megan Wilson, position -> Assistant Manager, phone -> 555-432-1098, email -> megan.wilson@example.com}]   |
-----------------------------------------------------------------------------------------------------------------------+
result2=hiveCtx.sql("Select * from jsontable where location='Los Angeles'")
result2.show()
```

```
    +---------------+-----------+--------------+-------+--------------------+
    |           name|   location|       address|revenue|           employees|
    +---------------+-----------+--------------+-------+--------------------+
    |Branch Office 2|Los Angeles|456 Oak Avenue|    5.5|[{name -> Michael...|
    |Branch Office 4|Los Angeles| 101 Pine Lane|    8.1|[{name -> Christo...|
    +---------------+-----------+--------------+-------+--------------------+
```

```
result3=hiveCtx.sql("Select * from jsontable where employees[0].name='John Doe'")
result3.show()
```

```
    +---------------+--------+---------------+-------+--------------------+
    |           name|location|        address|revenue|           employees|
    +---------------+--------+---------------+-------+--------------------+
    |Branch Office 1|New York|123 Main Street|    3.2|[{name -> John Do...|
    +---------------+--------+---------------+-------+--------------------+
```