In [1]:
```
pip install pyspark
```

```
Collecting pyspark
  Downloading pyspark-3.3.0.tar.gz (281.3 MB)
Collecting py4j==0.10.9.5
  Downloading py4j-0.10.9.5-py2.py3-none-any.whl (199 kB)
Building wheels for collected packages: pyspark
  Building wheel for pyspark (setup.py): started
  Building wheel for pyspark (setup.py): still running...
  Building wheel for pyspark (setup.py): still running...
  Building wheel for pyspark (setup.py): finished with status 'done'
  Created wheel for pyspark: filename=pyspark-3.3.0-py2.py3-none-any.whl size=281764040
sha256=4fa74a71414c6414e27c3820c3dd80ec48ddb443c56263d5734578874f2d76cf
  Stored in directory: c:\users\ajinkya\appdata\local\pip\cache\wheels\05\75\73\81f84d17
4299abca38dd6a06a5b98b08ae25fce50ab8986fa1
Successfully built pyspark
Installing collected packages: py4j, pyspark
Successfully installed py4j-0.10.9.5 pyspark-3.3.0
Note: you may need to restart the kernel to use updated packages.
```

In [2]:
```
import pyspark
```

In [30]:
```
import pandas as pd

display(pd.read_csv('test1.csv'))

type(pd.read_csv('test1.csv'))
```

	name	age
0	Ajinkya	32
1	Narendra	29
2	Amit	33
3	Nikhil	30

Out[30]:
```
pandas.core.frame.DataFrame
```

In [5]:
```
from pyspark.sql import SparkSession
```

In [6]:
```
spark = SparkSession.builder.appName('Practice').getOrCreate()
```

In [7]:
```
spark
```

Out[7]:

SparkSession - in-memory

SparkContext

Spark UI

Version	`v3.3.0`
Master	`local[*]`
AppName	`Practice`

In [9]:
```
spark.read.option('header', 'true').csv('test1.csv')
```

Out[9]:
```
DataFrame[name: string, age: string]
```

In [10]:
```
spark.read.option('header', 'true').csv('test1.csv').show()
```

```
+--------+---+
|    name|age|
+--------+---+
| Ajinkya| 32|
|Narendra| 29|
|    Amit| 33|
|  Nikhil| 30|
+--------+---+
```

In [11]:
```
df_pyspark = spark.read.option('header', 'true').csv('test1.csv')
type(df_pyspark)
```

Out[11]:
```
pyspark.sql.dataframe.DataFrame
```

In [12]:
```
df_pyspark.head(4)
```

Out[12]:
```
[Row(name='Ajinkya', age='32'),
 Row(name='Narendra', age='29'),
 Row(name='Amit', age='33'),
 Row(name='Nikhil', age='30')]
```

In [13]:
```
df_pyspark.printSchema()
```

```
root
 |-- name: string (nullable = true)
 |-- age: string (nullable = true)
```