Sql Session

import findspark  
findspark.init()  
from pyspark.sql import SparkSession  
spark= SparkSession.builder.master("local[\*]").getOrCreate()  
spark.conf.set("spark.sql.repl.eagerEval.enabled",True)

**Adding schema**

#Method 1

rdd= spark.sparkContext.parallelize([('Nikunj',69)])  
df=spark.createDataFrame(rdd,['Name','Age'])  
print(df.collect())

#method 2  
from pyspark.sql import SparkSession,Row

l=[('Nikunj',69)]  
rdd= spark.sparkContext.parallelize(l)  
Person=Row('name','age')  
person=rdd.map(lambda r: Person(\*r))  
df=spark.createDataFrame(person)  
print(df.collect())

#method3  
from pyspark.sql.types import \*  
l=[('Nikunj',69),('Jayu',25)]  
rdd= spark.sparkContext.parallelize(l)  
schema=StructType([  
 StructField('name',StringType(),True),  
 StructField('roll\_no',IntegerType(),False)])  
df=spark.createDataFrame(rdd,schema)  
print((df.collect()))

Dataframe

**cache**()

Persists the **[DataFrame](https://spark.apache.org/docs/2.1.0/api/python/pyspark.sql.html" \l "pyspark.sql.DataFrame" \o "pyspark.sql.DataFrame)** with the default storage level (**MEMORY\_AND\_DISK**).