*# Import the necessary module*from pyspark.sql import \*  
from pyspark.sql.types import \*  
  
*# Create a SparkSession object*spark = SparkSession.builder.appName('ReadJSON').getOrCreate()  
  
*#------------------------------------------------------------------------  
# Read a single json file into a DataFrame using .json() method and print its schema and data*df = spark.read.json("C:/Users/hars2071/Downloads/DEPractice/emp.json")  
df.printSchema()  
df.show()

*Output:*

root

|-- Id: long (nullable = true)

|-- Name: string (nullable = true)

|-- Salary: long (nullable = true)

+---+------+------+

| Id| Name|Salary|

+---+------+------+

| 1|Harsha| 20000|

| 2| Kiran| 25000|

| 3|Harika| 30000|

+---+------+------+

*#------------------------------------------------------------------------  
# Read a json file into a DataFrame and explicitly define the datatypes for columns*schema = StructType().add(field='Id',data\_type=IntegerType())\  
 .add(field='Name',data\_type=StringType())\  
 .add(field='Salary',data\_type=IntegerType())  
  
df1 = spark.read.json("C:/Users/hars2071/Downloads/DEPractice/emp.json",schema=schema)  
df1.printSchema()  
df1.show()

*Output:*

root

|-- Id: integer (nullable = true)

|-- Name: string (nullable = true)

|-- Salary: integer (nullable = true)

+---+------+------+

| Id| Name|Salary|

+---+------+------+

| 1|Harsha| 20000|

| 2| Kiran| 25000|

| 3|Harika| 30000|

+---+------+------+

*#------------------------------------------------------------------------  
# Read a multiLine json file into a DataFrame using .format() method*df2 = spark.read.format("json").option('multiLine',True).load("C:/Users/hars2071/Downloads/DEPractice/employeesML.json")  
  
df2.show()

*Output:*

+---+---------+---------+--------+------+

| Id|TitleName|firstName|lastName|salary|

+---+---------+---------+--------+------+

| 1|Developer| Harsha| Vardhan| 20000|

| 2|Developer| Harika| Nara| 30000|

| 3|Developer| Kiran| Kumar| 35000|

+---+---------+---------+--------+------+

*#------------------------------------------------------------------------  
# Read multiple JSON files into a DataFrame just by passing all file names with a separated comma as a path*df3 = spark.read.json(["C:/Users/hars2071/Downloads/DEPractice/emp.json","C:/Users/hars2071/Downloads/DEPractice/emp1.json"])  
  
df3.show()

*Output:*

+---+-------+------+

| Id| Name|Salary|

+---+-------+------+

| 4|Vardhan| 20000|

| 5| Venkat| 25000|

| 6|Krishna| 30000|

| 1| Harsha| 20000|

| 2| Kiran| 25000|

| 3| Harika| 30000|

+---+-------+------+  
  
*#------------------------------------------------------------------------  
# Read all JSON files into a DataFrame by passing the wildcard syntax*df4 = spark.read.json("C:/Users/hars2071/Downloads/DEPractice/\*.json")  
df4.printSchema()  
df4.show()