*# import the necessary libraries*from pyspark.sql import \*  
  
*# create a new SparkSession with a specified application name*spark = SparkSession.builder.appName('ReadCSVFile').getOrCreate()

*#------------------------------------------------------------------------*  
*# Read a single CSV file into a DataFrame using .CSV() method and print its schema and data*df = spark.read.csv("C:/Users/hars2071/Downloads/DEPractice/cust1.csv", header=True)  
df.printSchema() *# print the schema of the DataFrame*df.show() *# print the data of the DataFrame*

*Output:*

root

|-- Id: string (nullable = true)

|-- Name: string (nullable = true)

|-- Location : string (nullable = true)

|-- Salary: string (nullable = true)

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

| Id| Name|Location |Salary|

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

| 1|Harsha| BLR| 10000|

| 2|Harish| HYD| 20000|

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

*#------------------------------------------------------------------------  
# Read a single CSV file into a DataFrame using .format() method and print its schema and data*df1 = spark.read.format('csv')\  
 .option('header',True)\  
 .option('inferSchema',True)\  
 .load("C:/Users/hars2071/Downloads/DEPractice/cust1.csv")  
df1.printSchema() *# print the schema of the DataFrame*df1.show() *# print the data of the DataFrame*

*OutPut:*

root

|-- Id: integer (nullable = true)

|-- Name: string (nullable = true)

|-- Location : string (nullable = true)

|-- Salary: integer (nullable = true)

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

| Id| Name|Location |Salary|

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

| 1| Harsha| BLR| 10000|

| 2| Harish| HYD| 20000|

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

*#------------------------------------------------------------------------  
# Read multiple CSV files into a single DataFrame using .load() method and print its schema and data*df2 = spark.read.format('csv')\  
 .option('header',True)\  
 .option('inferSchema',True)\  
 .load(["C:/Users/hars2071/Downloads/DEPractice/cust1.csv",  
 "C:/Users/hars2071/Downloads/DEPractice/cust2.csv"])  
df2.printSchema() *# print the schema of the DataFrame*df2.show() *# print the data of the DataFrame*

*OutPut:*

root

|-- Id: integer (nullable = true)

|-- Name: string (nullable = true)

|-- Location : string (nullable = true)

|-- Salary: integer (nullable = true)

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

| Id| Name|Location |Salary|

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

| 1|Harsha| BLR| 10000|

| 2|Harish| HYD| 20000|

| 3|Venkat| HYD| 20000|

| 4|Harika| HYD| 25000|

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

*#------------------------------------------------------------------------  
# Read all CSV files in a directory into a single DataFrame using .load() method and print its schema and data*df3 = spark.read.format('csv')\  
 .option('header',True)\  
 .option('inferSchema',True)\  
 .load("C:/Users/hars2071/Downloads/DEPractice/")  
df3.printSchema() *# print the schema of the DataFrame*df3.show() *# print the data of the DataFrame*

*OutPut:*

root

|-- Id: integer (nullable = true)

|-- Name: string (nullable = true)

|-- Location : string (nullable = true)

|-- Salary: integer (nullable = true)

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

| Id| Name|Location |Salary|

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

| 1| Harsha| BLR| 10000|

| 2| Harish| HYD| 20000|

| 3| Venkat| HYD| 20000|

| 4| Harika| HYD| 25000|

| 5|Krishna| BLR| 20000|

| 6| Kiran| HYD| 25000|

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