E1

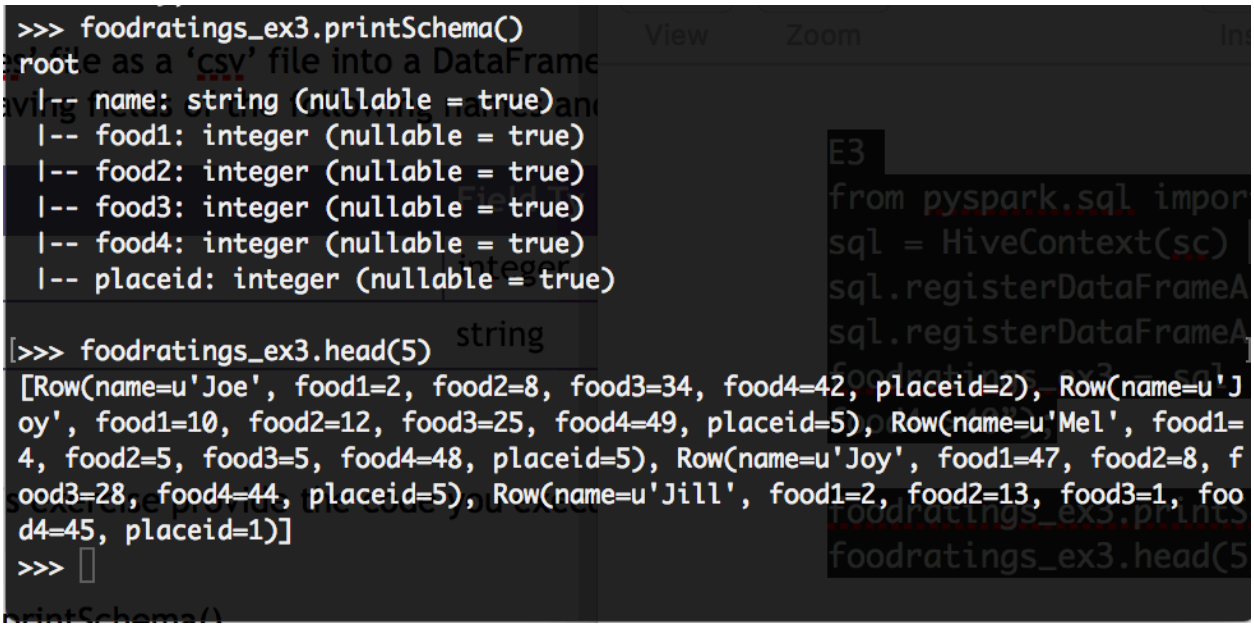
```
from pyspark.sql.types import *

foodratingsStruct = StructType(
    [
        StructField("name", StringType(), True),
        StructField("food1",IntegerType(), True),
        StructField("food2",IntegerType(), True),
        StructField("food3",IntegerType(), True),
        StructField("food4",IntegerType(), True),
        StructField("placeid",IntegerType(), True)
    ]
)

ex1_foodratings = spark.read.schema(foodratingsStruct).csv("hdfs:/user/
maria_dev/foodratings101106.txt")

ex1_foodratings.printSchema()
ex1_foodratings.head(5)
```

```
>>> ex1_foodratings.printSchema()
root
 |-- name: string (nullable = true)
 |-- food1: integer (nullable = true)
 |-- food2: integer (nullable = true)
 |-- food3: integer (nullable = true)
 |-- food4: integer (nullable = true)
 |-- placeid: integer (nullable = true)

>>> ex1_foodratings.head(5)
[Row(name=u'Joy', food1=11, food2=2, food3=29, food4=6, placeid=2), Row(name=u'J
oe', food1=2, food2=8, food3=34, food4=42, placeid=2), Row(name=u'Mel', food1=33
, food2=25, food3=46, food4=5, placeid=3), Row(name=u'Jill', food1=20, food2=45,
 food3=4, food4=3, placeid=2), Row(name=u'Joe', food1=44, food2=22, food3=10, fo
od4=20, placeid=3)]
>>>
```

```
E2
from pyspark.sql.types import *

foodplacesStruct = StructType(
    [
        StructField("placeid", IntegerType(), True),
        StructField("placename",StringType(), True),
    ]
)

ex2_foodplaces = spark.read.schema(foodplacesStruct).csv("hdfs:/user/
maria_dev/foodplaces101106.txt")
ex2_foodplaces.printSchema()
ex2_foodplaces.head(5)
```

```
>>> ex2_foodplaces = spark.read.schema(foodplacesStruct).csv("hdfs:/user/maria_d
ev/foodplaces101106.txt")
>>> ex2_foodplaces.printSchema()
root
 |-- placeid: integer (nullable = true)
 |-- placename: string (nullable = true)

>>> ex2_foodplaces.head(5)
[Row(placeid=1, placename=u'China Bistro'), Row(placeid=2, placename=u'Atlantic'
), Row(placeid=3, placename=u'Food Town'), Row(placeid=4, placename=u"Jake's"),
Row(placeid=5, placename=u'Soup Bowl')]
>>>
```

```
E3
from pyspark.sql import *
sql = HiveContext(sc)
sql.registerDataFrameAsTable(ex1_foodratings,"foodratingsT")
sql.registerDataFrameAsTable(ex2_foodplaces,"foodplacesT")
foodratings_ex3 = sql.sql("select * from foodratingsT where food2 < 25 and
food4 >40”);

foodratings_ex3.printSchema()
foodratings_ex3.head(5)
```



```
foodplaces_ex3 = sql.sql("select * from foodplacesT where placeid>3")
foodplaces_ex3.printSchema()
foodplaces_ex3.head(5)
```

```
>>> foodplaces_ex3 = sql.sql("select * from foodplacesT where placeid>3")
>>> foodplaces_ex3.printSchema()
root
 |-- placeid: integer (nullable = true)
 |-- placename: string (nullable = true)

>>> foodplaces_ex3.head(5)
[Row(placeid=4, placename=u"Jake's"), Row(placeid=5, placename=u'Soup Bowl')]
>>>
```

Ex4

```
foodratings_ex4 = ex1_foodratings.filter(ex1_foodratings['name']=="Mel")
foodratings_ex4 = ex1_foodratings.filter(ex1_foodratings['food3']<25)
foodratings_ex4.printSchema()
foodratings_ex4.head(5)
```

```
>>> foodratings_ex4.printSchema()
root
 |-- name: string (nullable = true)
 |-- food1: integer (nullable = true)
 |-- food2: integer (nullable = true)
 |-- food3: integer (nullable = true)
 |-- food4: integer (nullable = true)
 |-- placeid: integer (nullable = true)

>>> foodratings_ex4.head(5)
[Row(name=u'Jill', food1=20, food2=45, food3=4, food4=3, placeid=2), Row(name=u'
Joe', food1=44, food2=22, food3=10, food4=20, placeid=3), Row(name=u'Jill', food
1=30, food2=49, food3=8, food4=46, placeid=1), Row(name=u'Joe', food1=12, food2=
21, food3=7, food4=17, placeid=1), Row(name=u'Mel', food1=12, food2=37, food3=10
, food4=38, placeid=2)]
>>>
```

EX5

```
foodratings_ex5 = ex1_foodratings.select('name','placeid')
foodratings_ex5.printSchema()
foodratings_ex5.head(5)
```

```
>>> foodratings_ex5.printSchema()
root
 |-- name: string (nullable = true)
 |-- placeid: integer (nullable = true)

>>> foodratings_ex5.head(5)
[Row(name=u'Joy', placeid=2), Row(name=u'Joe', placeid=2), Row(name=u'Mel', plac
eid=3), Row(name=u'Jill', placeid=2), Row(name=u'Joe', placeid=3)]
>>>
```

EX6

```
condition = [ex1_foodratings.placeid==ex2_foodplaces.placeid]
ex6 =ex1_foodratings.join(ex2_foodplaces, condition, 'inner')
ex6.printSchema()
ex6.head(5)
```

```
>>> ex6.printSchema()
root
 |-- name: string (nullable = true)
 |-- food1: integer (nullable = true)
 |-- food2: integer (nullable = true)
 |-- food3: integer (nullable = true)
 |-- food4: integer (nullable = true)
 |-- placeid: integer (nullable = true)
 |-- placeid: integer (nullable = true)
 |-- placename: string (nullable = true)

>>> ex6.head(5)
[Row(name=u'Joy', food1=11, food2=2, food3=29, food4=6, placeid=2, placeid=2, pl
acename=u'Atlantic'), Row(name=u'Joe', food1=2, food2=8, food3=34, food4=42, pla
ceid=2, placeid=2, placename=u'Atlantic'), Row(name=u'Mel', food1=33, food2=25,
food3=46, food4=5, placeid=3, placeid=3, placename=u'Food Town'), Row(name=u'Jil
l', food1=20, food2=45, food3=4, food4=3, placeid=2, placeid=2, placename=u'Atla
ntic'), Row(name=u'Joe', food1=44, food2=22, food3=10, food4=20, placeid=3, plac
eid=3, placename=u'Food Town')]
>>>
```