

Milestone2

Group6

Jinyang Ruan, Rusu Wu, Yiyao, Brian Chan, Junqiao Mou

Part 1:Data preparation

1.

- a:** airports.dat (1.08 MB)
airlines.dat (388 KB)
routes.dat(2.27 MB)
- b:** airports: 14110tuples, 14 attributes
airlines:21317 tuples, 8 attributes
routes:67663 tuples, 9attributes

2.

- a. File format: “.dat”**
Delimiter: “,”

Part2:

Screenshot Group 1:

```

In [2]: from pyspark.sql import SparkSession
from sedona.register import SedonaRegister, SedonaSerializer
from sedona.utils import SedonaKryoRegistrar, KryoSerializer
import geopandas as gpd

spark = SparkSession. \
    builder. \
    appName('Python Spark Apache Sedona example'). \
    master('local[*]'). \
    config("spark.serializer", KryoSerializer.getName()). \
    config("spark.kryo.registrator", SedonaKryoRegistrar.getName()). \
    config("spark.jars.packages", \
        'org.apache.sedona:sedona-python-adapter-3.0.2.12:1.0.1-incubating,' \
        'org.datasyslab:geotools-wrapper:geotools-24.1'). \
    getOrCreate()

SedonaRegistrar.registerAll(spark)

21/10/10 12:07:00 WARN Utils: Your hostname, rusu-virtual-machine resolves to a loopback address: 127.0.1.1; using 1
92.168.42.128 instead (on interface ens33)
21/10/10 12:07:00 WARN Utils: Set SPARK LOCAL IP if you need to bind to another address
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/home/rusu/spark-3.0.3-bin-hadoop3.2/ja
rs/spark-unsafe-2.12-3.0.3.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
Ivy Default Cache set to: /home/rusu/.ivy2/cache
The jars for the packages stored in: /home/rusu/.ivy2/jars
:: loading settings :: url = jar:file:/home/rusu/spark-3.0.3-bin-hadoop3.2/jars/ivy-2.4.0.jar!/org/apache/ivy/core/s
ettings/ivysettings.xml
org.apache.sedona:sedona-python-adapter-3.0.2.12 added as a dependency
org.datasyslab:geotools-wrapper added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent-a0e59408-3f7d-4c00-a6ef-9ee99cac0f97;1.0
confs: [default]
found org.apache.sedona:sedona-python-adapter-3.0.2.12:1.0.1-incubating in central
found org.locationtech.jts#jts-core:1.18.0 in central
found org.wololo#jts2geojson:0.16.1 in central
found com.fasterxml.jackson.core#jackson-databind:2.12.2 in central
found com.fasterxml.jackson.core#jackson-annotations:2.12.2 in central
found com.fasterxml.jackson.core#jackson-core:2.12.2 in central
found org.apache.sedona:sedona-core-3.0.2.12:1.0.1-incubating in central
found org.apache.sedona:sedona-sql-3.0.2.12:1.0.1-incubating in central
found org.datasyslab:geotools-wrapper:geotools-24.1 in central
:: resolution report :: resolve 861ms :: artifacts dl 28ms
:: modules in use:
com.fasterxml.jackson.core#jackson-annotations:2.12.2 from central in [default]
com.fasterxml.jackson.core#jackson-core:2.12.2 from central in [default]
com.fasterxml.jackson.core#jackson-databind:2.12.2 from central in [default]
org.apache.sedona:sedona-core-3.0.2.12:1.0.1-incubating from central in [default]
org.apache.sedona:sedona-python-adapter-3.0.2.12:1.0.1-incubating from central in [default]
org.apache.sedona:sedona-sql-3.0.2.12:1.0.1-incubating from central in [default]
org.datasyslab:geotools-wrapper:geotools-24.1 from central in [default]
org.locationtech.jts#jts-core:1.18.0 from central in [default]
org.wololo#jts2geojson:0.16.1 from central in [default]
:: evicted modules:
org.locationtech.jts#jts-core:1.18.1 by [org.locationtech.jts#jts-core:1.18.0] in [default]
-----
| modules | | artifacts | | | | |
|---|---|---|---|---|---|---|
| conf | number | search | dwnlded | evicted | number | dwnlded |
|-----|
| default | 10 | 0 | 0 | 1 | 9 | 0 |
-----
:: retrieving :: org.apache.spark#spark-submit-parent-a0e59408-3f7d-4c00-a6ef-9ee99cac0f97
confs: [default]
9 artifacts copied, 9 already retrieved (0KB/70ms)
21/10/10 12:07:03 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).

```

Out[21]: True

```
In [3]: states_wkt = spark.read.option("delimiter", "\t").option("header", "false").csv("boundary-each-state.tsv").toDF("s_name", "s_bound")
states_wkt.show()
states_wkt.printSchema()

states = states_wkt.selectExpr("s_name", "ST_GeomFromWKT(s_bound) as s_bound")
states.show()
states.printSchema()
states.createOrReplaceTempView("states")
```

```
+-----+-----+
|   s_name|   s_bound|
+-----+-----+
|   Alaska|POLYGON((-141.020...|
|  Alabama|POLYGON((-88.1955...|
| Arkansas|POLYGON((-94.0416...|
|  Arizona|POLYGON((-112.598...|
| California|POLYGON((-124.400...|
|  Colorado|POLYGON((-109.044...|
| Connecticut|POLYGON((-73.4875...|
|  Delaware|POLYGON((-75.7919...|
|  Florida|POLYGON((-87.6050...|
|  Georgia|POLYGON((-85.6082...|
|   Hawaii|POLYGON((-154.628...|
|    Iowa|POLYGON((-95.7623...|
|   Idaho|POLYGON((-117.031...|
| Illinois|POLYGON((-90.6290...|
|  Indiana|POLYGON((-87.5253...|
|   Kansas|POLYGON((-102.050...|
| Kentucky|POLYGON((-89.5372...|
| Louisiana|POLYGON((-94.0430...|
|Massachusetts|POLYGON((-72.7789...|
|   Maryland|POLYGON((-79.4778...|
+-----+-----+
only showing top 20 rows
```

```
root
|-- s_name: string (nullable = true)
|-- s_bound: string (nullable = true)
```

```
+-----+-----+
|   s_name|   s_bound|
+-----+-----+
|   Alaska|POLYGON((-141.02...|
|  Alabama|POLYGON((-88.195...|
| Arkansas|POLYGON((-94.041...|
|  Arizona|POLYGON((-112.59...|
| California|POLYGON((-124.40...|
|  Colorado|POLYGON((-109.04...|
| Connecticut|POLYGON((-73.487...|
|  Delaware|POLYGON((-75.791...|
|  Florida|POLYGON((-87.605...|
|  Georgia|POLYGON((-85.608...|
|   Hawaii|POLYGON((-154.62...|
|    Iowa|POLYGON((-95.762...|
|   Idaho|POLYGON((-117.03...|
| Illinois|POLYGON((-90.629...|
|  Indiana|POLYGON((-87.525...|
|   Kansas|POLYGON((-102.05...|
| Kentucky|POLYGON((-89.537...|
| Louisiana|POLYGON((-94.043...|
|Massachusetts|POLYGON((-72.778...|
|   Maryland|POLYGON((-79.477...|
+-----+-----+
only showing top 20 rows
```

```
root
|-- s_name: string (nullable = true)
|-- s_bound: geometry (nullable = false)
```

```
In [4]: cities_csv = spark.read.option("delimiter", ",").option("header", "false").csv("cities.csv").toDF("c_name", "c_loc")
cities_csv.show()
cities_csv.printSchema()

cities = cities_csv.selectExpr("c_name", "ST_PointFromText(c_loc, '\') as c_loc")
cities.show()
cities.printSchema()
cities.createOrReplaceTempView("cities")

+-----+-----+
| c_name|      c_loc|
+-----+-----+
|Pullman|-117.167126 46.73...|
|Phoenix|-112.092128 33.50...|
+-----+-----+

root
 |-- c_name: string (nullable = true)
 |-- c_loc: string (nullable = true)

+-----+-----+
| c_name|      c_loc|
+-----+-----+
|Pullman|POINT (-117.16712...|
|Phoenix|POINT (-112.09212...|
+-----+-----+

root
 |-- c_name: string (nullable = true)
 |-- c_loc: geometry (nullable = false)
```

```
In [5]: city_per_state = spark.sql("select * from states s, cities c where ST_Contains(s.s_bound, c.c_loc)")
city_per_state.show()

21/10/10 12:08:34 WARN JoinQuery: UseIndex is true, but no index exists. Will build index on the fly.

+-----+-----+-----+-----+
| s_name| s_bound| c_name|      c_loc|
+-----+-----+-----+-----+
|Washington|POLYGON ((-123.32...|Pullman|POINT (-117.16712...|
| Arizona|POLYGON ((-112.59...|Phoenix|POINT (-112.09212...|
+-----+-----+-----+-----+

In [6]: dist_to_seattle = spark.sql("select c_name, ST_Distance(c_loc, ST_Point(-122.313323, 47.622715)) as dist from cities")
dist_to_seattle.show()

+-----+-----+
| c_name|      dist|
+-----+-----+
|Pullman| 5.22191398200746|
|Phoenix|17.428931546401486|
+-----+-----+
```

```
In [7]: geopandas_df = gpd.GeoDataFrame(states.toPandas(), geometry="s_bound")
        geopandas_df
```

Out[7]:

	s_name	s_bound
0	Alaska	POLYGON ((-141.02050 70.01870, -141.72910 70.1...
1	Alabama	POLYGON ((-88.19550 35.00410, -85.60680 34.991...
2	Arkansas	POLYGON ((-94.04160 33.02250, -91.20570 33.007...
3	Arizona	POLYGON ((-112.59890 36.99930, -110.86300 37.0...
4	California	POLYGON ((-124.40090 41.99830, -123.62370 42.0...
5	Colorado	POLYGON ((-109.04480 37.00040, -102.04240 36.9...
6	Connecticut	POLYGON ((-73.48750 42.04980, -73.42470 42.051...
7	Delaware	POLYGON ((-75.79190 39.71880, -75.78370 39.521...
8	Florida	POLYGON ((-87.60500 30.99880, -86.56130 30.996...
9	Georgia	POLYGON ((-85.60820 34.99740, -84.72660 34.990...
10	Hawaii	POLYGON ((-154.62840 19.51580, -154.82350 19.8...
11	Iowa	POLYGON ((-95.76230 40.58480, -93.54450 40.578...
12	Idaho	POLYGON ((-117.03190 49.00000, -116.04860 49.0...
13	Illinois	POLYGON ((-90.62900 42.51160, -87.02130 42.492...
14	Indiana	POLYGON ((-87.52530 41.76110, -84.80900 41.761...
15	Kansas	POLYGON ((-102.05060 40.00340, -102.05060 40.0...
16	Kentucky	POLYGON ((-89.53720 36.49860, -89.30100 36.507...
17	Louisiana	POLYGON ((-94.04300 33.02250, -93.00480 33.017...
18	Massachusetts	POLYGON ((-72.77890 42.00030, -72.74050 42.033...
19	Maryland	POLYGON ((-79.47780 39.72200, -78.36000 39.722...
20	Maine	POLYGON ((-71.01290 45.34250, -70.95250 45.332...
21	Michigan	POLYGON ((-88.37130 48.30330, -87.60500 48.010...
22	Minnesota	POLYGON ((-96.45170 43.50080, -91.21950 43.501...
23	Missouri	POLYGON ((-91.74680 40.61810, -91.68090 40.559...
24	Mississippi	POLYGON ((-90.30490 35.00410, -88.19550 35.007...
25	Montana	POLYGON ((-111.05120 44.47500, -111.36840 44.7...
26	North Carolina	POLYGON ((-78.48500 33.79630, -79.67420 34.803...
27	North Dakota	POLYGON ((-104.05010 48.99550, -104.04460 45.9...
28	Nebraska	POLYGON ((-104.05430 42.99860, -104.05430 41.0...
29	New Hampshire	POLYGON ((-71.50450 45.01530, -71.49770 45.053...
30	New Jersey	POLYGON ((-74.69500 41.35720, -74.65590 41.339...
31	New Mexico	POLYGON ((-109.04480 36.99710, -109.04890 31.3...
32	Nevada	POLYGON ((-120.00370 41.99220, -117.96020 42.0...
33	New York	POLYGON ((-79.76240 42.51420, -79.06720 42.778...
34	Ohio	POLYGON ((-82.34250 38.43850, -82.29170 38.570...
35	Oklahoma	POLYGON ((-103.00510 37.00150, -103.00370 36.5...
36	Oregon	POLYGON ((-124.24440 46.25870, -124.03840 46.2...
37	Pennsylvania	POLYGON ((-80.51740 42.32610, -80.08210 42.396...
38	Rhode Island	POLYGON ((-71.05680 41.37060, -71.05410 41.370...
39	South Carolina	POLYGON ((-80.70010 32.04880, -80.89780 32.045...
40	South Dakota	POLYGON ((-104.05010 45.94350, -104.05290 43.0...
41	Tennessee	POLYGON ((-88.06780 36.67390, -87.85220 36.635...
42	Texas	POLYGON ((-106.57150 31.86590, -106.50420 31.7...
43	Utah	POLYGON ((-114.04910 36.99820, -109.04620 37.0...
44	Virginia	POLYGON ((-83.67530 36.60170, -83.56130 36.651...
45	Vermont	POLYGON ((-73.34270 45.01040, -73.18270 45.013...
46	Washington	POLYGON ((-123.32080 49.00230, -123.03380 49.0...
47	Wisconsin	POLYGON ((-90.63030 42.51160, -87.02410 42.495...
48	West Virginia	POLYGON ((-79.47650 39.71880, -79.48060 39.205...
49	Wyoming	POLYGON ((-104.05560 41.00370, -104.05840 44.9...

Screenshot Group 2:

Since there is no geometry data in the data set, the output only can be string.

```
In [1]: from pyspark.sql import SparkSession
from sedona.register import SedonaRegistrator
from sedona.utils import SedonaKryoRegistrator, KryoSerializer
import geopandas as gpd

spark = SparkSession. \
    builder. \
    appName('Python Spark Apache Sedona'). \
    master('spark://rusu-virtual-machine:7077'). \
    config("spark.serializer", KryoSerializer.getName()). \
    config("spark.kryo.registrator", SedonaKryoRegistrator.getName()). \
    config('spark.jars.packages',
           'org.apache.sedona:sedona-python-adapter-3.0_2.12:1.0.1-incubating,'
           'org.datasyslab:geotools-wrapper:geotools-24.1'). \
    getOrCreate()

SedonaRegistrator.registerAll(spark)
```

21/10/10 14:57:42 WARN Utils: Your hostname, rusu-virtual-machine resolves to a loopback address: 127.0.1.1; using 192.168.42.128 instead (on interface ens33)
21/10/10 14:57:42 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/home/rusu/spark-3.0.3-bin-hadoop3.2/jars/spark-unsafe_2.12-3.0.3.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
Ivy Default Cache set to: /home/rusu/.ivy2/cache
The jars for the packages stored in: /home/rusu/.ivy2/jars
:: loading settings :: url = jar:file:/home/rusu/spark-3.0.3-bin-hadoop3.2/jars/ivy-2.4.0.jar!/org/apache/ivy/core/settings/ivysettings.xml
org.apache.sedona#sedona-python-adapter-3.0_2.12 added as a dependency
org.datasyslab#geotools-wrapper added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent-ae2ca91d-683d-4ad5-b0ed-ca63fce27cf1;1.0
 confs: [default]
 found org.apache.sedona#sedona-python-adapter-3.0_2.12;1.0.1-incubating in central
 found org.locationtech.jts#jts-core;1.18.0 in central
 found org.wololo#jts2geojson;0.16.1 in central
 found com.fasterxml.jackson.core#jackson-databind;2.12.2 in central
 found com.fasterxml.jackson.core#jackson-annotations;2.12.2 in central
 found com.fasterxml.jackson.core#jackson-core;2.12.2 in central
 found org.apache.sedona#sedona-core-3.0_2.12;1.0.1-incubating in central
 found org.apache.sedona#sedona-sql-3.0_2.12;1.0.1-incubating in central
 found org.datasyslab#geotools-wrapper;geotools-24.1 in central
:: resolution report :: resolve 952ms :: artifacts dl 45ms
 found org.datasyslab#geotools-wrapper;geotools-24.1 in central
:: resolution report :: resolve 952ms :: artifacts dl 45ms
 :: modules in use:
 com.fasterxml.jackson.core#jackson-annotations;2.12.2 from central in [default]
 com.fasterxml.jackson.core#jackson-core;2.12.2 from central in [default]
 com.fasterxml.jackson.core#jackson-databind;2.12.2 from central in [default]
 org.apache.sedona#sedona-core-3.0_2.12;1.0.1-incubating from central in [default]
 org.apache.sedona#sedona-python-adapter-3.0_2.12;1.0.1-incubating from central in [default]
 org.apache.sedona#sedona-sql-3.0_2.12;1.0.1-incubating from central in [default]
 org.datasyslab#geotools-wrapper;geotools-24.1 from central in [default]
 org.locationtech.jts#jts-core;1.18.0 from central in [default]
 org.wololo#jts2geojson;0.16.1 from central in [default]
 :: evicted modules:
 org.locationtech.jts#jts-core;1.18.1 by [org.locationtech.jts#jts-core;1.18.0] in [default]
 |-----|
 | conf | | modules | | artifacts |
 |-----| |-----| |-----| |-----|
 | number | search | dwnlded | evicted | | number | dwnlded |
 |-----| |-----| |-----| |-----|
 | default | 10 | 0 | 0 | 1 | | 9 | 0 |
 |-----| |-----| |-----| |-----|
 :: retrieving :: org.apache.spark#spark-submit-parent-ae2ca91d-683d-4ad5-b0ed-ca63fce27cf1
 confs: [default]
 0 artifacts copied, 9 already retrieved (0kB/46ms)
21/10/10 14:57:45 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
21/10/10 14:57:50 WARN SparkContext: Please ensure that the number of slots available on your executors is limited b
y the number of cores to task cpus and not another custom resource. If cores is not the limiting resource then dynam
ic allocation will not work properly!

Out[1]: True

```
In [58]: airlines_dat = spark.read.option("delimiter", ";").option("header", "false").csv("airlines.dat").toDF("AirlineID", "Name", "Alias", "ITAT", "ICAO", "Callsign", "Country", "Active")
airlines_dat.show()
airlines_dat.printSchema()

airlines = airlines_dat.selectExpr("AirlineID", "Name", "Alias", "ITAT", "ICAO", "Callsign", "Country", "Active")
airlines.show()
airlines.printSchema()
airlines.createOrReplaceTempView("airlines")
```

AirlineID	Name	Alias	ITAT	ICAO	Callsign	Country	Active
-1	Unknown	\N	-	N/A	\N	\N	Y
1	Private flight	\N	-	N/A	null	null	Y
2	135 Airways	\N	null	GNL	GENERAL	United States	N
3	1Time Airline	\N	1T	RNX	NEXTIME	South Africa	Y
4	2 Sqn No 1 Elemen...	\N	null	WYT	null	United Kingdom	N
5	213 Flight Unit	\N	null	TFU	null	Russia	N
6	223 Flight Unit S...	\N	null	CHD	CHKALOVSK-AVIA	Russia	N
7	224th Flight Unit	\N	null	TTF	CARGO UNIT	Russia	N
8	247 Jet Ltd	\N	null	TWF	CLOUD RUNNER	United Kingdom	N
9	3D Aviation	\N	null	SEC	SECUREX	United States	N
10	40-Mile Air	\N	Q5	MLA	MILE-AIR	United States	Y
11	4D Air	\N	null	QRT	QUARTET	Thailand	N
12	611897 Alberta Li...	\N	null	THD	DONUT	Canada	N
13	Ansett Australia	\N	AN	AAA	ANSETT	Australia	Y
14	Abacus International	\N	1B	null	null	Singapore	Y
15	Abelag Aviation	\N	W9	AAB	ABG	Belgium	N
16	Army Air Corps	\N	null	AAC	ARMYAIR	United Kingdom	N
17	Aero Aviation Cen...	\N	null	AAD	SUNRISE	Canada	N
18	Aero Servicios Ej...	\N	null	SII	ASEISA	Mexico	N
19	Aero Biniza	\N	null	BZS	BINIZA	Mexico	N

only showing top 20 rows

```
root
-- AirlineID: string (nullable = true)
-- Name: string (nullable = true)
-- Alias: string (nullable = true)
-- ITAT: string (nullable = true)
-- ICAO: string (nullable = true)
-- Callsign: string (nullable = true)
-- Country: string (nullable = true)
-- Active: string (nullable = true)
```

AirlineID	Name	Alias	ITAT	ICAO	Callsign	Country	Active
-1	Unknown	\N	-	N/A	\N	\N	Y
1	Private flight	\N	-	N/A	null	null	Y
2	135 Airways	\N	null	GNL	GENERAL	United States	N
3	1Time Airline	\N	1T	RNX	NEXTIME	South Africa	Y
4	2 Sqn No 1 Elemen...	\N	null	WYT	null	United Kingdom	N
5	213 Flight Unit	\N	null	TFU	null	Russia	N
6	223 Flight Unit S...	\N	null	CHD	CHKALOVSK-AVIA	Russia	N
7	224th Flight Unit	\N	null	TTF	CARGO UNIT	Russia	N
8	247 Jet Ltd	\N	null	TWF	CLOUD RUNNER	United Kingdom	N
9	3D Aviation	\N	null	SEC	SECUREX	United States	N
10	40-Mile Air	\N	Q5	MLA	MILE-AIR	United States	Y
11	4D Air	\N	null	QRT	QUARTET	Thailand	N
12	611897 Alberta Li...	\N	null	THD	DONUT	Canada	N
13	Ansett Australia	\N	AN	AAA	ANSETT	Australia	Y
14	Abacus International	\N	1B	null	null	Singapore	Y
15	Abelag Aviation	\N	W9	AAB	ABG	Belgium	N
16	Army Air Corps	\N	null	AAC	ARMYAIR	United Kingdom	N
17	Aero Aviation Cen...	\N	null	AAD	SUNRISE	Canada	N
18	Aero Servicios Ej...	\N	null	SII	ASEISA	Mexico	N
19	Aero Biniza	\N	null	BZS	BINIZA	Mexico	N

only showing top 20 rows

```
root
-- AirlineID: string (nullable = true)
-- Name: string (nullable = true)
-- Alias: string (nullable = true)
-- ITAT: string (nullable = true)
-- ICAO: string (nullable = true)
-- Callsign: string (nullable = true)
-- Country: string (nullable = true)
-- Active: string (nullable = true)
```



```
In [68]: airports_dat = spark.read.option("delimiter", ",").option("header", "false").csv("airports.dat").toDF("AirportID", "Name", "City", "Country", "IATA", "ICAO", "Latitude", "Longitude", "Altitude")
airports_dat.show()
airports_wkt.printSchema()

airports = airports_dat.selectExpr("AirportID", "Name", "City", "Country", "IATA", "ICAO", "Latitude", "Longitude", "Altitude")
airports.show()
airports.printSchema()
airports.createOrReplaceTempView("airports")
```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+										
[AirportID		Name		City	Country		IATA	ICAO	Latitude	
titude		Timezone	DST	Tz	Type	Source			Longitude	
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+										
5282	1	Goroka Airport		Goroka	Papua New Guinea		GKA	AYGA	-6.081689834590001	
	10	U Pacific/Port_Moresby		airport	OurAirports				145.391998291	
20	2	Madang Airport		Madang	Papua New Guinea		MAG	AYMD	-5.20707988739	
	10	U Pacific/Port_Moresby		airport	OurAirports				145.789001465	
5388	3	Mount Hagen Kagam...		Mount Hagen	Papua New Guinea		HGU	AYMH	-5.826789855957031	
	10	U Pacific/Port_Moresby		airport	OurAirports				144.29600524902344	
239	4	Nadzab Airport		Nadzab	Papua New Guinea		LAE	AYNZ	-6.569803	
	10	U Pacific/Port_Moresby		airport	OurAirports				146.725977	
146	5	Port Moresby Jack...		Port Moresby	Papua New Guinea		POM	AYPY	-9.443380355834961	
	10	U Pacific/Port_Moresby		airport	OurAirports				147.22000122070312	
19	6	Wewak Internation...		Wewak	Papua New Guinea		WWK	AYWK	-3.58383011818	
	10	U Pacific/Port_Moresby		airport	OurAirports				143.669006348	
	7	Narsarsuaq Airport		Narsarsuaq	Greenland		UAK	BGBW	61.1604995728	
									-45.4259986877	

112	7	Narsarsuaq Airport	Narssarsuaq	Greenland	UAK BGBW	61.1604995728	-45.4259986877
	-3	E	America/Godthab airport OurAirports				
283	8	Godthaab / Nuuk A...	Godthaab	Greenland	GOH BGGH	64.19090271	-51.6781005859
	-3	E	America/Godthab airport OurAirports				
165	9	Kangerlussuaq Air...	Søndrestrom	Greenland	SFJ BGSF	67.0122218992	-50.7116031647
	-3	E	America/Godthab airport OurAirports				
251	10	Thule Air Base	Thule	Greenland	THU BGTL	76.5311965942	-68.7032012939
	-4	E	America/Thule airport OurAirports				
6	11	Akureyri Airport	Akureyri	Iceland	AEY BIAR	65.66000366210938	-18.07270050048828
	0	N	Atlantic/Reykjavik airport OurAirports				
76	12	Egilsstaðir Airport	Egilsstaðir	Iceland	EGS BIEG	65.2833023071289	-14.401399612426758
	0	N	Atlantic/Reykjavik airport OurAirports				
24	13	Hornafjörður Airport	Hofn	Iceland	HFN BIHN	64.295601	-15.2272
	0	N	Atlantic/Reykjavik airport OurAirports				
48	14	Húsavík Airport	Húsavík	Iceland	HZK BIHU	65.952301	-17.426001
	0	N	Atlantic/Reykjavik airport OurAirports				
8	15	Ísafjörður Airport	Ísafjörður	Iceland	IFJ BIIS	66.05809783935547	-23.135299682617188
	0	N	Atlantic/Reykjavik airport OurAirports				
	16	Keflavík Internat...	Keflavík	Iceland	KEF BIKF	63.985000610352	-22.605600357056

171	16	Keflavik Internat...	Keflavik	Iceland	KEF BIKF	63.985000610352	-22.605600357056
	0	N Atlantic/Reykjavik	airport	OurAirports			
11	17	Patreksfjörður Ai...	Patreksfjörður	Iceland	PFJ BIPA	65.555801	-23.965
	0	N Atlantic/Reykjavik	airport	OurAirports			
48	18	Reykjavik Airport	Reykjavik	Iceland	RKV BIRK	64.1299972534	-21.9405994415
	0	N Atlantic/Reykjavik	airport	OurAirports			
10	19	Siglufjörður Airport	Siglufjörður	Iceland	SIJ BISI	66.133301	-18.9167
	0	N Atlantic/Reykjavik	airport	OurAirports			
326	20	Vestmannaeyjar Ai...	Vestmannaeyjar	Iceland	VEY BIVM	63.42430114746094	-20.278900146484375
	0	N Atlantic/Reykjavik	airport	OurAirports			

```
-----+-----+-----+-----
only showing top 20 rows
```

```
root
|-- AirportID: string (nullable = true)
|-- Name: string (nullable = true)
|-- City: string (nullable = true)
|-- Country: string (nullable = true)
```

```
root
|-- AirportID: string (nullable = true)
|-- Name: string (nullable = true)
|-- City: string (nullable = true)
|-- Country: string (nullable = true)
|-- IATA: string (nullable = true)
|-- ICAO: string (nullable = true)
|-- Latitude: string (nullable = true)
|-- Longitude: string (nullable = true)
|-- Altitude: string (nullable = true)
|-- Timezone: string (nullable = true)
|-- DST: string (nullable = true)
|-- Tz: string (nullable = true)
|-- Type: string (nullable = true)
|-- Source: string (nullable = true)
```


	AirportID	Name	City	Country	IATA	ICAO	Latitude	Longitude	Altitude
	Timezone	DST	Tz	Type	Source				
	1	Goroka Airport	Goroka	Papua New Guinea	GKA	AYGA	-6.081689834590001	145.391998291	
5282	10	U Pacific/Port Moresby airport OurAirports							
	2	Madang Airport	Madang	Papua New Guinea	MAG	AYMD	-5.20707988739	145.789001465	
20	10	U Pacific/Port Moresby airport OurAirports							
	3	Mount Hagen Kagam...	Mount Hagen	Papua New Guinea	HGU	AYMH	-5.826789855957031	144.29600524902344	
5388	10	U Pacific/Port Moresby airport OurAirports							
	4	Nadzab Airport	Nadzab	Papua New Guinea	LAE	AYNZ	-6.569803	146.725977	
239	10	U Pacific/Port Moresby airport OurAirports							
	5	Port Moresby Jack...	Port Moresby	Papua New Guinea	POM	AYPY	-9.443380355834961	147.22000122070312	
146	10	U Pacific/Port Moresby airport OurAirports							
	6	Wewak Internation...	Wewak	Papua New Guinea	WWK	AYWK	-3.58383011818	143.669006348	
19	10	U Pacific/Port Moresby airport OurAirports							
	7	Narsarsuaq Airport	Narsarsuaq	Greenland	UAK	BGBW	61.1604995728	-45.4259986877	

112	-3	E America/Godthab airport OurAirports							
	8	Godthaab / Nuuk A...	Godthaab	Greenland	GOH	BGGH	64.19090271	-51.6781005859	
283	-3	E America/Godthab airport OurAirports							
	9	Kangerlussuaq Air...	Sondrestrom	Greenland	SFJ	BGSF	67.0122218992	-50.7116031647	
165	-3	E America/Godthab airport OurAirports							
	10	Thule Air Base	Thule	Greenland	THU	BGTL	76.5311965942	-68.7032012939	
251	-4	E America/Thule airport OurAirports							
	11	Akureyri Airport	Akureyri	Iceland	AEY	BIAR	65.66000366210938	-18.07270050048828	
6	0	N Atlantic/Reykjavik airport OurAirports							
	12	Egilsstaðir Airport	Egilsstaðir	Iceland	EGS	BIEG	65.2833023071289	-14.401399612426758	
76	0	N Atlantic/Reykjavik airport OurAirports							
	13	Hornafjörður Airport	Hofn	Iceland	HFN	BIHN	64.295601	-15.2272	
24	0	N Atlantic/Reykjavik airport OurAirports							
	14	Húsavík Airport	Husavik	Iceland	HZK	BIHU	65.952301	-17.426001	
48	0	N Atlantic/Reykjavik airport OurAirports							
	15	Isafjörður Airport	Isafjordur	Iceland	IFJ	BIIS	66.05809783935547	-23.135299682617188	
8	0	N Atlantic/Reykjavik airport OurAirports							
	16	Keflavik Internat...	Keflavik	Iceland	KEF	BIKF	63.985000610352	-22.605600357056	

171	0	N Atlantic/Reykjavik airport OurAirports							
	17	Patreksfjörður Ai...	Patreksfjordur	Iceland	PFJ	BIPA	65.555801	-23.965	
11	0	N Atlantic/Reykjavik airport OurAirports							
	18	Reykjavik Airport	Reykjavik	Iceland	RKV	BIRK	64.1299972534	-21.9405994415	
48	0	N Atlantic/Reykjavik airport OurAirports							
	19	Siglufjörður Airport	Siglufjordur	Iceland	SIJ	BISI	66.133301	-18.9167	
10	0	N Atlantic/Reykjavik airport OurAirports							
	20	Vestmannaeyjar Ai...	Vestmannaeyjar	Iceland	VEY	BIVM	63.42430114746094	-20.278900146484375	
326	0	N Atlantic/Reykjavik airport OurAirports							

only showing top 20 rows

```

root
|-- AirportID: string (nullable = true)
|-- Name: string (nullable = true)
|-- City: string (nullable = true)
|-- Country: string (nullable = true)

```

```

|-- Latitude: string (nullable = true)
|-- Longitude: string (nullable = true)
|-- Altitude: string (nullable = true)
|-- Timezone: string (nullable = true)
|-- DST: string (nullable = true)
|-- Tz: string (nullable = true)
|-- Type: string (nullable = true)
|-- Source: string (nullable = true)

```

```
In [69]: routes_dat = spark.read.option("delimiter",",").option("header", "false").csv("routes.dat").toDF("Airline","AirlineID",
routes_dat.show()
routes_dat.printSchema()

routes = routes_dat.selectExpr("Airline","AirlineID","SourceaAirport","SourceAirportID","DestinationAirport","Destina
routes.show()
routes.printSchema()
routes.createOrReplaceTempView("routes")
```

Airline	AirlineID	SourceaAirport	SourceAirportID	DestinationAirport	DestinationAirportID	Codeshare	Stops	Equipment
2B	410	AER	2965	KZN	2990	null	0	CR2
2B	410	ASF	2966	KZN	2990	null	0	CR2
2B	410	ASF	2966	MRV	2962	null	0	CR2
2B	410	CEK	2968	KZN	2990	null	0	CR2
2B	410	CEK	2968	OVB	4078	null	0	CR2
2B	410	DME	4029	KZN	2990	null	0	CR2
2B	410	DME	4029	NBC	6969	null	0	CR2
2B	410	DME	4029	TGK	\N	null	0	CR2
2B	410	DME	4029	UUA	6160	null	0	CR2
2B	410	EGO	6156	KGD	2952	null	0	CR2
2B	410	EGO	6156	KZN	2990	null	0	CR2
2B	410	GYD	2922	NBC	6969	null	0	CR2
2B	410	KGD	2952	EGO	6156	null	0	CR2
2B	410	KZN	2990	AER	2965	null	0	CR2
2B	410	KZN	2990	ASF	2966	null	0	CR2
2B	410	KZN	2990	CEK	2968	null	0	CR2
2B	410	KZN	2990	DME	4029	null	0	CR2
2B	410	KZN	2990	EGO	6156	null	0	CR2
2B	410	KZN	2990	LED	2948	null	0	CR2
2B	410	KZN	2990	SVX	2975	null	0	CR2

only showing top 20 rows

```
root
|-- Airline: string (nullable = true)
|-- AirlineID: string (nullable = true)
|-- SourceaAirport: string (nullable = true)
|-- SourceAirportID: string (nullable = true)
|-- DestinationAirport: string (nullable = true)
|-- DestinationAirportID: string (nullable = true)
|-- Codeshare: string (nullable = true)
|-- Stops: string (nullable = true)
|-- Equipment: string (nullable = true)
```


Airline	AirlineID	SourceaAirport	SourceAirportID	DestinationAirport	DestinationAirportID	Codeshare	Stops	Equipment
2B	410	AER	2965	KZN	2990	null	0	CR2
2B	410	ASF	2966	KZN	2990	null	0	CR2
2B	410	ASF	2966	MRV	2962	null	0	CR2
2B	410	CEK	2968	KZN	2990	null	0	CR2
2B	410	CEK	2968	OVB	4078	null	0	CR2
2B	410	DME	4029	KZN	2990	null	0	CR2
2B	410	DME	4029	NBC	6969	null	0	CR2
2B	410	DME	4029	TGK	\N	null	0	CR2
2B	410	DME	4029	UUA	6160	null	0	CR2
2B	410	EGO	6156	KGD	2952	null	0	CR2
2B	410	EGO	6156	KZN	2990	null	0	CR2
2B	410	GYD	2922	NBC	6969	null	0	CR2
2B	410	KGD	2952	EGO	6156	null	0	CR2
2B	410	KZN	2990	AER	2965	null	0	CR2
2B	410	KZN	2990	ASF	2966	null	0	CR2
2B	410	KZN	2990	CEK	2968	null	0	CR2
2B	410	KZN	2990	DME	4029	null	0	CR2
2B	410	KZN	2990	EGO	6156	null	0	CR2
2B	410	KZN	2990	LED	2948	null	0	CR2
2B	410	KZN	2990	SVX	2975	null	0	CR2

only showing top 20 rows

```
root
|-- Airline: string (nullable = true)
|-- AirlineID: string (nullable = true)
|-- SourceaAirport: string (nullable = true)
|-- SourceAirportID: string (nullable = true)
|-- DestinationAirport: string (nullable = true)
|-- DestinationAirportID: string (nullable = true)
|-- Codeshare: string (nullable = true)
|-- Stops: string (nullable = true)
|-- Equipment: string (nullable = true)
```

In []:

Screenshot Group 3:



3.0.3

Spark Master at spark://rusu-virtual-machine:7077

URL: spark://rusu-virtual-machine:7077

Alive Workers: 1

Cores in use: 2 Total, 2 Used

Memory in use: 2.8 GiB Total, 1024.0 MB Used

Resources in use:

Applications: 1 Running, 6 Completed

Drivers: 0 Running, 0 Completed

Status: ALIVE

Workers (1)

Worker Id	Address	State	Cores	Memory	Resources
worker-20210909075809-192.168.42.128-41287	192.168.42.128:41287	ALIVE	2 (2 Used)	2.8 GiB (1024.0 MB Used)	

Running Applications (1)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
app-20211010145751-0006	(kill) Python Spark Apache Sedona	2	1024.0 MB		2021/10/10 14:57:51	rusu	RUNNING	2.5 h

Completed Applications (6)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
app-20211010130538-0005	Python Spark Apache Sedona	2	1024.0 MB		2021/10/10 13:05:38	rusu	FINISHED	1.8 h
app-20211010130311-0004	Python Spark Apache Sedona example	2	1024.0 MB		2021/10/10 13:03:11	rusu	FINISHED	1.6 min
app-20211010113106-0003	Python Spark Apache Sedona example	2	1024.0 MB		2021/10/10 11:31:06	rusu	FINISHED	7.8 min
app-20211010105508-0002	Python Spark SQL basic example	2	1024.0 MB		2021/10/10 10:55:08	rusu	FINISHED	1.2 min
app-20210913215315-0001	Python Spark SQL basic example	2	1024.0 MB		2021/09/13 21:53:15	rusu	FINISHED	636.5 h
app-20210913215202-0000	Python Spark SQL basic example	2	1024.0 MB		2021/09/13 21:52:02	rusu	FINISHED	43 s