Running neo4j on docker:

a) Loading data from CSV:

```
LOAD CSV FROM
"https://raw.githubusercontent.com/jpatokal/openflights/master/data/airports.dat" AS row
WITH toInteger(row[0]) AS Id,
    row[1] AS Name,
    row[2] AS City,
    row[3] AS Country,
    row[4] AS Code,
    row[5] AS ICAO,
    toFloat(row[6]) AS Latitude,
    toFloat(row[7]) AS Longitude,
    toFloat(row[8]) AS Altitude,
    toInteger(row[9]) AS Timezone
MERGE (a:Airport {Id: Id})
  ON CREATE SET a.Name = Name,
        a.City = City,
        a.Country = Country,
        a.Code = Code,
        a.ICAO = ICAO,
        a.Latitude = Latitude,
        a.Longitude = Longitude,
        a.Altitude = Altitude,
        a.Timezone = Timezone


LOAD CSV FROM
"https://raw.githubusercontent.com/jpatokal/openflights/master/data/airlines.dat" AS row
WITH toInteger(row[0]) AS Id,
```

```
    row[1] AS Name,
    row[2] AS Alias,
    row[3] AS IATA,
    row[4] AS ICAO,
    row[5] AS Callsign,
    row[6] AS Country,
    row[7] AS Active
MERGE (a:Airline {Id: Id})
  ON CREATE SET a.Name = Name,
          a.Alias = Alias,
          a.IATA = IATA,
          a.ICAO = ICAO,
          a.Callsign = Callsign,
          a.Country = Country,
          a.Active = Active



LOAD CSV FROM
"https://raw.githubusercontent.com/jpatokal/openflights/master/data/routes.dat" AS row
CREATE (:Route (AirlineCode: row[0], AirlineId : toInteger(row[1]), SourceAirport : row[2],
    SourceAirportId : toInteger(row[3]), DestinationAirport : row[4], DestinationAirportId :
toInteger(row[5]),
    Codeshare: row[6],Stops : toInteger(row[7]), Equipment: row[8])
```
Renaming attributes was not needed as the attributes names are already named according the referenced website during load from csv.
Creating relations:
```
:auto USING PERIODIC COMMIT
LOAD CSV  FROM
"https://raw.githubusercontent.com/jpatokal/openflights/master/data/routes.dat" as row
MATCH (A:Airline) WHERE A.Id = toInteger(row[1])
MATCH (R:Route) WHERE R.AirlineId = toInteger(row[1])
CREATE (R)-[:OPERATED_BY]->(A)



:auto USING PERIODIC COMMIT 1000 LOAD CSV  FROM
"https://raw.githubusercontent.com/jpatokal/openflights/master/data/routes.dat" as row
MATCH (A:Airport) WHERE A.Id = toInteger(row[3])
MATCH (R:Route) WHERE R.SourceAirportId = toInteger(row[3])
CREATE (R)-[:STARTS_FROM]->(A)

:auto USING PERIODIC COMMIT 1000
```

```
LOAD CSV  FROM
"https://raw.githubusercontent.com/jpatokal/openflights/master/data/routes.dat" as row
MATCH (A:Airport) WHERE A.Id = toInteger(row[5])
MATCH (R:Route) WHERE R.DestinationAirportId = toInteger(row[5])
CREATE (R)-[:END_TO]->(A)
```

b) Answer to the 8 queries:

a. MATCH (A:Airport) WHERE A.Country = 'Canada' RETURN COUNT(A)

b.

```
MATCH (R:Route), (A:Airport) WHERE R.SourceAirportId = A.Id
AND A.Name = 'St. Anthony Airport'
RETURN COUNT(R)
```

MATCH (R:Route), (A:Airport), (A1:Airport)
WHERE R.SourceAirportId = A.Id
AND R.DestinationAirportId = A1.Id
AND A.Name = 'St. Anthony Airport'
AND A1.Country = 'Canada'
RETURN R, A, A1

neo4j$ MATCH (R:Route), (A:Airport), (A1:Airport) WHERE R.SourceAirportId = A.Id AND R.DestinationAirportId = A1.Id AND A.Name =...

R

```
{
    "AirlineCode": "PB",
    "SourceAirportId": 24,
    "Equipment": "SFB",
    "SourceAirport": "YAY",
    "Stops": 0,
    "DestinationAirportId": 5485,
    "DestinationAirport": "YBX"
}
```

A

```
{
    "Timezone": -3,
    "Country": "Canada",
    "Latitude": 51.3918991089,
    "ICAO": "CYAY",
    "City": "St. Anthony",
    "Id": 24,
    "Longitude": -56.083099365200006,
    "Code": "YAY",
    "Altitude": 108.0,
    "Name": "St. Anthony Airport"
}
```

A1

```
{
    "Timezone": -4,
    "Country": "Canada",
    "Latitude": 51.443599700899995,
    "ICAO": "CYBX",
    "City": "Lourdes-De-Blanc-
Sablon",
    "Id": 5485,
    "Code": "YBX",
    "Longitude": -57.185298919699996,
    "Altitude": 121.0,
    "Name": "Lourdes de Blanc Sablon
Airport"
}
```

Started streaming 3 records after 1 ms and completed after 68 ms.

c.
MATCH (AR:Airline), (R:Route), (A:Airport), (A1:Airport)
WHERE AR.Id = R.AirlineId
AND R.SourceAirportId = A.Id
AND R.DestinationAirportId = A1.Id
AND (A.NAME = 'St. Anthony Airport' OR A1.Name = 'St. Anthony Airport')
RETURN DISTINCT AR.Name

```
1 MATCH (AR:Airline), (R:Route), (A:Airport), (A1:Airport)
2 WHERE AR.Id = R.AirlineId
3 AND R.SourceAirportId = A.Id
4 AND R.DestinationAirportId = A1.Id
5 AND (A.NAME = 'St. Anthony Airport' OR A1.Name = 'St. Anthony Airport')
6 RETURN DISTINCT AR.Name
```

neo4j$ MATCH (AR:Airline), (R:Route), (A:Airport), (A1:Airport) WHERE AR.Id = R.Airli

Table

(no changes, no records)

Code

d.

`MATCH (R:Route) RETURN COUNT(R) AS NumberOfRoutes`

neo4j$ **MATCH (R:Route) RETURN COUNT(R) AS NumberOfRoutes**

neo4j$ MATCH (R:Route) RETURN COUNT(R) AS NumberOfRoutes

Table

NumberOfRoutes

A   67663
Text

MATCH (R:Route) RETURN (COUNT(R) - COUNT(R.Codeshare)) / toFloat(COUNT(R)) * 100 AS PercentageOfNotSharingFlight

```
neo4j$  MATCH (R:Route) RETURN (COUNT( R) - COUNT(R.Codeshare)) / toFloat(COUNT( R)) * 100 AS PercentageOfNotSharingFlight
```

To enjoy the full Neo4j Browser experience, we advise you to use   **Neo4j Browser Sync**

```
neo4j$ MATCH (R:Route) RETURN (COUNT( R) - COUNT(R.Codeshare)) / toFloat(COUNT( R)) * 100 AS PercentageOfNotSharingFlight
```

PercentageOfNotSharingFlight

78.42690983255251

e.

```
MATCH (R:Route), (A:Airline)
WHERE R.AirlineId = A.Id
WITH A.Name AS AirlineId, COUNT(*) AS NumberOfRoutes
WHERE  NumberOfRoutes > 1200
RETURN AirlineId, NumberOfRoutes
```

```
1  MATCH (R:Route), (A:Airline)
2  WHERE R.AirlineId = A.Id
3  WITH A.Name AS AirlineId, COUNT(*) AS NumberOfRoutes
4  WHERE  NumberOfRoutes > 1200
5  RETURN AirlineId, NumberOfRoutes
```

To enjoy the full Neo4j Browser experience, we advise you to use   **Neo4j Browser Sync**

```
neo4j$ MATCH (R:Route), (A:Airline) WHERE R.AirlineId = A.Id WITH A.Name AS AirlineId, COUNT(*) AS NumberOfRoutes WHERE NumberOfRoutes > 1200 RETURN AirlineId, …
```

AirlineId	NumberOfRoutes
"American Airlines"	2354
"Air China"	1260
"China Southern Airlines"	1454
"Delta Air Lines"	1981
"Ryanair"	2484
"China Eastern Airlines"	1263
"United Airlines"	2180
"US Airways"	1960

Started streaming 8 records in less than 1 ms and completed after 330 ms.

f.

```
CALL  {MATCH (A:Airport), (R:Route)
WHERE A.Id = R.DestinationAirportId
WITH A.Name AS AirportName, COUNT(*) AS NumberOfAirportAsDestination
WHERE NumberOfAirportAsDestination > 180
RETURN AirportName, NumberOfAirportAsDestination }
RETURN COUNT(AirportName) AS NumberOfAirportWithMoreThan80TimesAsDestination
```

```
1  CALL
2  {MATCH (A:Airport), (R:Route)
3  WHERE A.Id = R.DestinationAirportId
4  WITH A.Name AS AirportName, COUNT(*) AS NumberOfAirportAsDestination
5  WHERE NumberOfAirportAsDestination > 180
6  RETURN AirportName, NumberOfAirportAsDestination
7  }
```

neo4j$ CALL {MATCH (A:Airport), (R:Route) WHERE A.Id = R.DestinationAirportId WITH A.

NumberOfAirportWithMoreThan80TimesAsDestination

86

```
CALL  {MATCH (A:Airport), (R:Route)
WHERE A.Id = R.SourceAirportId
WITH A.Name AS AirportName, COUNT(*) AS NumberOfAirportAsSource
WHERE NumberOfAirportAsSource > 180
RETURN AirportName, NumberOfAirportAsSource }
RETURN COUNT(AirportName) AS NumberOfAirportWithMoreThan80TimesAsSource
```

```
1  CALL
2  {MATCH (A:Airport), (R:Route)
3  WHERE A.Id = R.SourceAirportId
4  WITH A.Name AS AirportName, COUNT(*) AS NumberOfAirportAsSource
5  WHERE NumberOfAirportAsSource > 180
6  RETURN AirportName, NumberOfAirportAsSource
7  }
```

neo4j$ CALL {MATCH (A:Airport), (R:Route) WHERE A.Id = R.SourceAirportId WITH A.N

NumberOfAirportWithMoreThan80TimesAsSource

84

The numbers are not equal.

g.

MATCH (A:Airport), (R:Route), (B:Airport) WHERE A.Id = R.SourceAirportId AND
R.DestinationAirportId = B.Id AND A.Name = 'St. Anthony Airport'
RETURN COUNT(DISTINCT B.Name)

```
⚠ 1  MATCH (A:Airport), (R:Route), (B:Airport)
  2  WHERE A.Id = R.SourceAirportId
  3  AND R.DestinationAirportId = B.Id
  4  AND A.Name = 'St. Anthony Airport'
  5  RETURN COUNT(DISTINCT B.Name)
```

To enjoy the full Neo4j Browser experience, we advise you to use   **Neo4j Browser Sync**

neo4j$ MATCH (A:Airport), (R:Route), (B:Airport) WHERE A.Id = R

Table

COUNT(DISTINCT B.Name)

Text

3

h.

MATCH (A:Airline), (R:Route), (AR:Airport)
WHERE A.Id = R.AirlineId AND R.DestinationAirportId = AR.Id
AND A.Name = 'Air France'
WITH AR.Country AS Country, COUNT(DISTINCT R.DestinationAirportId) AS NumberOfAirports
ORDER BY NumberOfAirports DESC
RETURN Country, NumberOfAirports

```
⚠ 1 MATCH (A:Airline), (R:Route), (AR:Airport)
  2 WHERE A.Id = R.AirlineId AND R.DestinationAirportId = AR.Id
  3 AND A.Name = 'Air France'
  4 WITH AR.Country AS Country, COUNT(DISTINCT R.DestinationAirportId) AS NumberOfAirports
  5 ORDER BY NumberOfAirports DESC
  6 RETURN Country, NumberOfAirports
```

`neo4j$ MATCH (A:Airline), (R:Route), (AR:Airport) WHERE A.Id = R.AirlineId AND R.DestinationAirportId = AR.Id AND A.Name = 'Air France'`

Country	NumberOfAirports
"United States"	141
"France"	29
"United Kingdom"	11
"China"	11
"Germany"	10
"Italy"	10
"Spain"	8
"Brazil"	6
"Mexico"	5

Started streaming 121 records after 1 ms and completed after 208 ms.

c) Two more queries:

a. Top ten countries based on number of airports:

MATCH (AR:Airport)

WITH AR.Country AS Country, COUNT(*) AS NumberOfAirports

ORDER BY NumberOfAirports DESC LIMIT 10

RETURN Country, NumberOfAirports

```
1 MATCH  (AR:Airport)
2 WITH AR.Country AS Country, COUNT(*) AS NumberOfAirports
3 ORDER BY NumberOfAirports DESC LIMIT 10
4 RETURN Country, NumberOfAirports
```

neo4j$ MATCH (AR:Airport) WITH AR.Country AS Country, COUNT(*) AS NumberOfAirports ORDER BY NumberOfAirports

Country	NumberOfAirports
"United States"	1512
"Canada"	430
"Australia"	334
"Brazil"	264
"Russia"	264
"Germany"	249
"China"	241
"France"	217
"United Kingdom"	167

Started streaming 10 records after 1 ms and completed after 190 ms.

b. Top ten airlines based on number of routes:

MATCH (AR:Airline), (R:Route)

WHERE AR.Id = R.AirlineId

WITH AR.Name AS Airline, COUNT(*) AS NumberOfRoutes

ORDER BY NumberOfRoutes DESC LIMIT 10

RETURN Airline, NumberOfRoutes

```
⚠ 1 MATCH  (AR:Airline), (R:Route)
  2 WHERE  AR.Id = R.AirlineId
```

```
Airline, COUNT(*) AS NumberOfRoutes
fRoutes DESC LIMIT 10
NumberOfRoutes
```

```
WHERE AR.Id = R.Airl
WITH AR.Name AS
ORDER BY NumberO
RETURN Airline,
```

e), (R:Route) WHERE AR.Id = R.AirlineId WITH AR.Name AS Airline, COUNT(*) AS NumberOfRoutes ORDER   neo4j$ MATCH (AR:Airlin

NumberOfRoutes		Airline
	Table	
2484	Text	"Ryanair"
2354		"American Airlines"
2180	Code	"United Airlines"
1981		"Delta Air Lines"
1960		"US Airways"
1454		"China Southern Airlines"
1263		"China Eastern Airlines"
1260		"Air China"
1146		"Southwest Airlines"