Exercise 3

The data root folder is dataset.

Due to computational limitations (both of the solution providers had laptops with only 4G ram, HDD and dual core mobile processors, 3 datafiles were imported fom all years uniformly (1 datafile per year).

File "Import Queries.txt" contains Cypher data import queries. "Analysis Queries.txt" contains data analysis queries. Queries are separated by — and question subparts.

APOC plugin has been used for date parsing / list utility functions.

Question 1: Import the Data

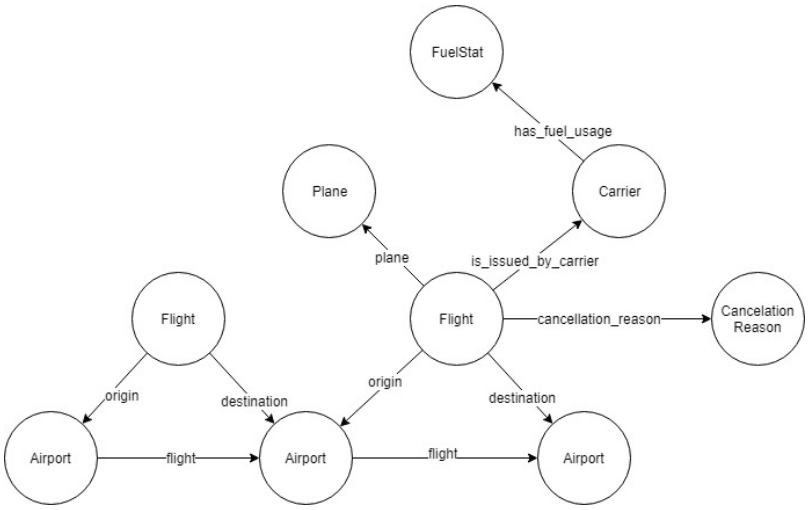

Figure 1: Partial schema of the problem.

The figure above represents the schema of the problem. We have decided to use nodes for flights, as flight queries require joining additional information on flight properties (e.g. getting plane engine type for a flight). Thus, all such joins have been extracted to separate nodes, and relationships have been added with flights.

Such operations could be executed without adding relationships. Though, relationships simplify queries and improve performance. These two arguments have been main motivators for establishing the schema above.

It should be noted that additional airport->flight->airport relationship has been added, because, currently, "Neo4j" shortestPath construct does not support 3 node search (when there's a partial node in the connection). It was a problem for the e) part of the data analysis section. We have also been thinking of removing flight nodes, and keeping flight relationships. Though, the queries would become cumbersome. Also, the engine does support hypergraphs (e.g. a set of airport, plane and carrier having flight relationships with another set of airport, plane and carrier). Such functionality would allow to discard flight nodes and query shortest paths, having only flight relationships.

Question 2: Data Analysis

a) 1) 2014-2015: yes (471949 to 503897). On 2016, they dropped to 460949.

2) Unique planes (based on tailnum) per year in air: 2014 (4399), 2015 (4347), 2016 (4362)

3) William B Hartsfield-Atlanta Intl (94201)

4) 1491

b) For 2014, 2015 and 2016, fuel consumption and cost correlates with number of flights.

Year	**Carrier**	**Gallons**	**USD**
2014	Delta Air Lines Inc.	3265645505	9321714459
2014	United Air Lines Inc.	3183378467	9042121024
2014	American Airlines Inc.	2473463889	6991774276
Year	**Carrier**	**Gallons**	**USD**
2015	Delta Air Lines Inc.	3388856207	7845050611
2015	United Air Lines Inc.	3215520613	5390875179
2015	American Airlines Inc.	3043664245	4955099986
Year	**Carrier**	**Gallons**	**USD**
2016	American Airlines Inc.	3595720988	4732547705
2016	Delta Air Lines Inc.	3412617055	5614257197
2016	United Air Lines Inc.	3261229135	4357641361

Year	Carrier	Flights
2014	Southwest Airlines Co.	91564
2014	Delta Air Lines Inc.	59030
2014	Atlantic Southeast Airlines	58508
2014	Skywest Airlines Inc.	50484
2014	American Airlines Inc.	45401
Year	**Carrier**	**Flights**
2015	Southwest Airlines Co.	109776
2015	Delta Air Lines Inc.	77255
2015	Skywest Airlines Inc.	50307
2015	Atlantic Southeast Airlines	49119
2015	United Air Lines Inc.	46084
Year	**Carrier**	**Flights**
2016	Southwest Airlines Co.	110806
2016	American Airlines Inc.	73802
2016	Delta Air Lines Inc.	71612
2016	Skywest Airlines Inc.	49628
2016	United Air Lines Inc.	45718

c) Distance covered and the number of flights per engine type per years.

Year	EngineType	Flights	Distance
"2014"	"4 Cycle"	526	496226
"2014"	"Electric"	72	81952
"2014"	"None"	248	162576
"2014"	"Reciprocating"	8890	7673440
"2014"	"Turbo-Fan"	729034	576156218
"2014"	"Turbo-Jet"	40148	36105294
"2014"	"Turbo-Prop"	17548	3934518
"2014"	"Turbo-Shaft"	1706	1490212

Figure 2: Distance covered and the number of flights per engine type 2014

Year	EngineType	Flights	Distance
"2015"	"2 Cycle"	146	367344
"2015"	"4 Cycle"	276	205582
"2015"	"None"	178	137682
"2015"	"Reciprocating"	5926	7364138
"2015"	"Turbo-Fan"	833380	691086072
"2015"	"Turbo-Jet"	45608	38679010
"2015"	"Turbo-Prop"	578	734658
"2015"	"Turbo-Shaft"	370	606368

Figure 3: Distance covered and the number of flights per engine type 2015

Year	EngineType	Flights	Distance
"2016"	"2 Cycle"	612	732676
"2016"	"4 Cycle"	3096	2445042
"2016"	"Electric"	1746	2016584
"2016"	"None"	214	284820
"2016"	"Reciprocating"	29804	31912248
"2016"	"Turbo-Fan"	768008	641999420
"2016"	"Turbo-Jet"	37492	32103128
"2016"	"Turbo-Prop"	6100	6208712
"2016"	"Turbo-Shaft"	2148	3100200

Figure 4: Distance covered and the number of flights per engine type 2016

d) 1) William B Hartsfield-Atlanta Intl (taxi In: 549519). William B Hartsfield-Atlanta Intl (taxi Out: 1515958)

2) 2014 - Chicago (4641, most frequent reason: National Air System), 2015 - Chicago (1456, most frequent reason: Carrier), 2016 - Chicago (1119, most frequent reason: Weather)

3) Worst - Southwest Airlines Co. (3172). Best - Hawaiian Airlines Inc. (20)

4) January (1). 19108 cancellations.

e) Shortest path - ADK -> ANC -> PHX -> DTW -> BGR

1) 4 times (counting hop count here)

2) Planes: "A320-232", "CL-600-2D24", "737-490", "757-26D" (counting distinct planes). Carriers: "Alaska Airlines Inc.", "Delta Air Lines Inc.", "Atlantic Southeast Airlines", "US Airways Inc." (counting distinct carriers).

3) 5470 km

4) 749 min = 12.483 h

5) Average delay for each flight (taken in and out delays into account): 45 min.