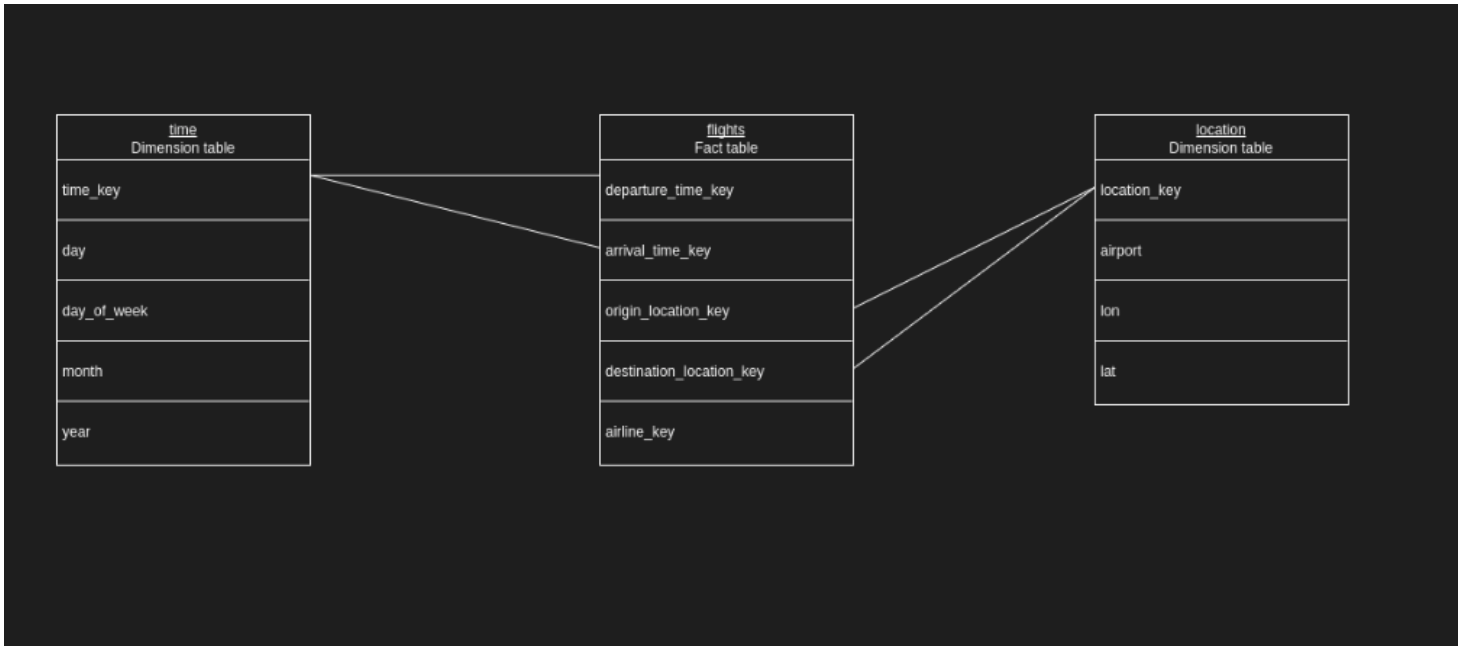
TDT4300 Datamining - Exercise 1

1 - Modeling



2 – OLAP operations

Report 1 - "The longest duration of any flight in the air"

Use MAX() on air_time to find the longest duration of any flight in the air.

Report 2 - "Average elapsed time for each airline company"

Drill down by moving down the concept hierarchy of the airline-dimension.

Report 3 - "The total number of flights flown in february "

Drill down by moving down the concept hierarchy of the time-dimension.

Report 4 - "Each month and the airport with the highest amount of arrival flights"

Drill down by moving down the concept hierarchy of the time-dimension and the location-dimension.

Report 5 - "Descending list of all months by the amount of total distance flown each month"

Drill down by moving down the concept hierarchy of the time-dimension.

4 - Multi-Dimensional Expressions (MDX)

Report 1 - "The longest duration of any flight in the air"

```
SELECT [Measures].[air time] ON COLUMNS
   FROM [cube]
```

	All
air time	690.0

Report 2 - "Average elapsed time for each airline company"

```
SELECT [Measures].[elapsed time] FROM Columns,
   [location].[airport] FROM ROWS
   FROM [cube]
```

	elapsed time
Lehigh Valley International Airport	729.0901497695852
Abilene Regional Airport	359.9032258064516
Albuquerque International Sunport	5588.0086405529955
Aberdeen Regional Airport	130.0014400921659
Southwest Georgia Regional Airport	142.26728110599078
Nantucket Memorial Airport	0.0
Waco Regional Airport	192.74395161290323
Arcata Airport	277.0552995391705
Atlantic City International Airport	1530.7744815668202
Adak Airport	45.540610599078335
Kodiak Airport	45.10397465437788
Alexandria International Airport	683.7975230414746
Augusta Regional Airport (Bush Field)	377.60167050691246
King Salmon Airport	0.0
Albany International Airport	2741.777649769585
Waterloo Regional Airport	130.0766129032258
Rick Husband Amarillo International Airport	1142.5861175115208
Ted Stevens Anchorage International Airport	6435.015264976959
Alpena County Regional Airport	101.82056451612902
Aspen-Pitkin County Airport	2205.9098502304146
Hartsfield-Jackson Atlanta International Airport	108067.93605990784
Appleton International Airport	628.7433755760369
Austin-Bergstrom International Airport	14445.018721198157
Asheville Regional Airport	469.7986751152074
Wilkes-Barre/Scranton International Airport	400.3856566820277
Kalamazoo/Battle Creek International Airport	355.1399769585254
Bradley International Airport	7559.383352534562
Bethel Airport	174.73444700460828
Meadows Field	920.5682603686636
Greater Binghamton Airport	33.74107142857143
Bangor International Airport	24.892857142857142
Birmingham-Shuttlesworth International Airport	3755.1241359447004
Billings Logan International Airport	655.8168202764977
Bismarck Municipal Airport	930.3683755760369
Bemidji Regional Airport	114.95046082949308
Bellingham International Airport	625.432603686636

See report_2.csv for full output.

Report 3 - "The total number of flights flown in February"

```
SELECT [Measures].[number of flights] FROM Columns,
   [time].[monthL] FROM ROWS
   FROM [cube]
```

	number of fligh...
> 1	469968
> 2	429191
> 3	504312
> 4	51298

Report 4 - “Each Month and the airport with the highest amount of arrival flights”

Note: we were unsure of why the distinct function did not affect the 'month' row, and could therefore not completely solve the problem. Did not manage to get in contact with the TA's in time.

```
SELECT
   [Measures].[number of flights] FROM columns,
   TopCount(
      {Distinct([time].[monthL].members)*[location].[airport].members},
      [Measures].[number of flights],
      [Measures].[number of flights]
   ) ON ROWS
FROM [cube]
```

		number of flights
3	Hartsfield-Jackson Atlanta International Airport	32754
1	Hartsfield-Jackson Atlanta International Airport	29512
2	Hartsfield-Jackson Atlanta International Airport	27366
3	Chicago O'Hare International Airport	26129
1	Chicago O'Hare International Airport	23484
3	Dallas/Fort Worth International Airport	23224
1	Dallas/Fort Worth International Airport	23153
2	Chicago O'Hare International Airport	21812
	Dallas/Fort Worth International Airport	20839
3	Denver International Airport	18609
	Los Angeles International Airport	18014
1	Los Angeles International Airport	17340
	Denver International Airport	17090
2	Los Angeles International Airport	15762
	Denver International Airport	15642
3	Phoenix Sky Harbor International Airport	14526
	George Bush Intercontinental Airport	14045
	San Francisco International Airport	13422
1	George Bush Intercontinental Airport	13376
	Phoenix Sky Harbor International Airport	13122
	San Francisco International Airport	12891
3	McCarran International Airport	12397
2	George Bush Intercontinental Airport	12231
	Phoenix Sky Harbor International Airport	11916
1	McCarran International Airport	11604
2	San Francisco International Airport	11569
3	Orlando International Airport	11296
2	McCarran International Airport	10570
3	Detroit Metropolitan Airport	10317
	LaGuardia Airport (Marine Air Terminal)	10140
1	Orlando International Airport	10083
3	Gen. Edward Lawrence Logan International Airport	9971
	Minneapolis-Saint Paul International Airport	9946
	Newark Liberty International Airport	9564
1	LaGuardia Airport (Marine Air Terminal)	9517
3	Salt Lake City International Airport	9476

See report_4.csv for full output.

Report 5 - "Descending list of all months by the amount of total distance flown each month"

```
SELECT
   [Measures].[total distance ] ON columns,
   TopCount(
      {[time].[monthL].members},
      [Measures].[total distance ],
      [Measures].[total distance ]
   ) ON ROWS
FROM [cube]
```

	total distance
3	4.11546494E8
1	3.77507097E8
2	3.43689908E8
4	4.1744774E7