TDT4300 — **Assignment 1**

DATA WAREHOUSING

tdt4300-undass@idi.ntnu.no

Spring 2022 SOLUTION

Problem

Download and unzip the file *data.zip*. Inside are the files *airlines.csv*, *airports.csv* and *flights.csv* displaying recorded data of flights, airlines and airports in USA, between 1st of January 2015 and 4th April 2015.

Check the source out to understand the meanings behind the columns. Data source: `https://www.kaggle.com/usdot/flight-delays?select=flights.csv`.

Make yourself familiar with the datasets. Your task is to provide the following reports:

Report 1 The longest duration of any flight in the air.
   MDX:

```
select [Measures].[Air_Time-MAX] on columns from [cube]
```

   Result:

	All
Air_Time-MAX	690

Report 2 Average elapsed time for each airline company.
   MDX:

```
select [Measures].[Elapsed_Time-AVG] on columns,
[Airline].[Code] on rows from [cube]
```

Result: Airline code or long name are fine for the solution.

	Elapsed_Time-AVG
AS	179.2018823068062
AA	171.6737130932138
US	151.47222675533433
DL	142.70279663637837
NK	158.61287054409004
UA	191.0209334133974
HA	101.54818110442292
B6	170.5659478117677
OO	99.74338521129704
EV	98.60321864971402
MQ	96.44852505328474
F9	154.0158112624071
WN	121.26467011651499
VX	208.2136130136

Report 3 The total number of flights flown in February.
   MDX:

```
select [Measures].[Flight Count] on columns from [cube]
where [StartTime].[Month].&[2015-02]
```

   Result:

	3. Feb 2015
Flight Count	429191

Report 4 Each Month and the airport with the highest amount of arrival flights.
   MDX:

```
select {[Measures].[Flight Count]} on columns,
    non empty [StartTime].[Month] *
    order(existing [ArrivalAirport].[Code],
    [Measures].[Flight Count], DESC).ITEM(0) on rows
from [cube]
```

Result: Airport code or long name are fine for the solution.

		Flight Count
2015 Jan	ATL	29492
2015 Feb	ATL	27366
2015 Mar	ATL	32775
2015 Apr	ATL	3317

Alternative Result: Airport code or long name are fine for the solution.

		Flight Count
2015 Jan	DFW	23173
2015 Feb	DFW	20839
2015 Mar	DFW	23208
2015 Apr	DFW	2333

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

MDX:

```
select [Measures].[Distance] on columns,
  order([StartTime].[Month], [Measures].[Distance], DESC) on rows
from [cube]
```

Result:

	Distance
2015 Mar	411546494
2015 Jan	377507097
2015 Feb	343689908
2015 Apr	41744774

Deliver the results AND your MDX queries See attachment for simple schema version.

To achieve this follow the instructions below.

1 Modeling

Given the dataset create a star schema compatible with the requirements above. Also define the concept hierarchies for each dimension. Briefly explain any assumptions you have made. We are primarily looking for you to show modeling principles for data warehousing.

1.1 Solution

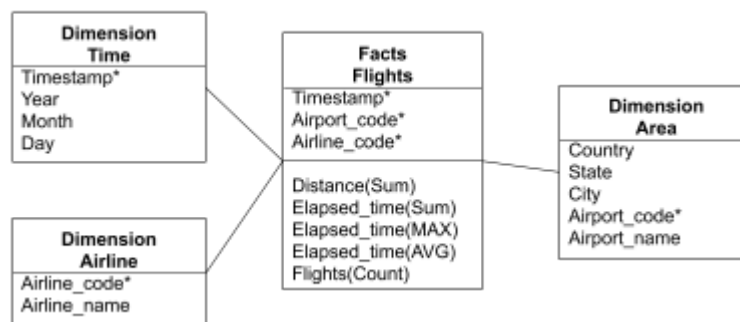Figure 1. Concept hierarkies (the v is an arrow pointing downwards)

Figure 2. Star schema

2 OLAP Operations

On top of your schema specify sequences of OLAP operations generating Report 1, Report 2 and Report 3. **Some may provide Report 4 instead of Report 1, which is OK**. Keep in mind following assumption: For all concept hierarchies assume the "All"-level (level with zero granularity) as default.

2.1 Solution

Report 1

- Drill down on no dimension and do for ALL, aggregate to the maximum air time of all fights.

Report 2

- Drill down on Airline dimension from ALL to Name, if the name is available in the main source table otherwise the code name has to be used.

Report 3

- Drill down on Time dimension from All to Day

- Slice on Time dimension with Day equal to 3

Report 4

- Drill down on Airport dimension from ALL to Name, if the name is available in the main source table otherwise the code name has to be used.

- Drill down on Time dimension from All to Day

3 Implementation of the Cube

You are going to use the icCube (an OLAP server with a web-based user interface) to implement the schema/cube you have created in the section 1. Download and install icCube on your computer. The provided dataset does not require any changes. All you need for creating a new schema/cube is the Builder tab. There you need to create a new schema, set the data source, define the dimensions and concept hierarchies, and define the cube and the measures.