Norwegian University of Science and Technology

TDT4300 — **Assignment 1**

DATA WAREHOUSING

tdt4300-undass@idi.ntnu.no

Spring 2022

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.

Report 2 Average elapsed time for each airline company.

Report 3 The total number of flights flown in February.

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

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

Deliver the results AND your MDX queries

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.

2 OLAP Operations

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

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[1] 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.

4 Multi-Dimensional Expressions (MDX)

Deploy your schema and switch to the MDX tab. Finally, define and execute MDX queries generating the requested reports.

[1] `http://www.iccube.com/download/`

Submission Requirements

In this assignment we expect you to submit following artifacts:

- A PDF file reporting the mdx code and the results of the sections 1, 2 and 4.
  - **Text must not be handwritten.** (Hand drawn figures are acceptable, but not wanted.)
  - Make sure that the document follows the usual conventions (**names**, assignment/task number, etc.).
- A file (**.icc-schema*) with the exported schema.

All assignment artifacts are to be delivered using ***BlackBoard***. You are allowed to **work in pairs**, however, the identical artifacts must be delivered individually.