Assignment 1: Data Warehousing
TDT4300

Jonas Myrlund

February 14, 2014

1

a) Explain in your own words the concepts of OLTP (Online Transaction Processing) and OLAP (Online Analytical Processing). Emphasize on the differences between the two concepts in terms of properties and usage.

OLTP, as the name implies, is used in typical transactional services. In practice, this means many users performing relatively simple updates that require high availability, concurrency, and efficiency. Doing bank transactions, updating customer relations etc. are examples of tasks that involve OLTP.

OLAP is more related to analyzing data. It usually involves a lower volume of transactions, and the queries are often quite complex and involving aggregations. Effectiveness can vary, but it is not unusual for a batch job to take several hours to run.

Moreover, the purposes of the systems differ in themselves: OLTP systems are involved with running the core business, while OLAP systems provide information that is typically used in planning and problem solving activities.

b) Explain the concept of data cube and the meaning of the term "cuboids".

An n-dimensional data cube is the generic n-dimensional equivalent of a typical 2-dimensional data table. It embodies the concept of extracting axes from a data set and handling it as a cube of n dimensions.

A cuboid is a possible n-dimensional cube that can be extracted from a larger data cube. For an n-dimensional data cube there are 2^n possible cuboids, ie. the number of possible subsets of the set of dimensions.

c) Explain the data cube operations slice, dice, rollup and drill-down.

Slice

Picking a subcube by selecting only records sharing one specific value for a chosen dimension.

Dice

Slicing with multiple possible values for the chosen dimension.

Rollup

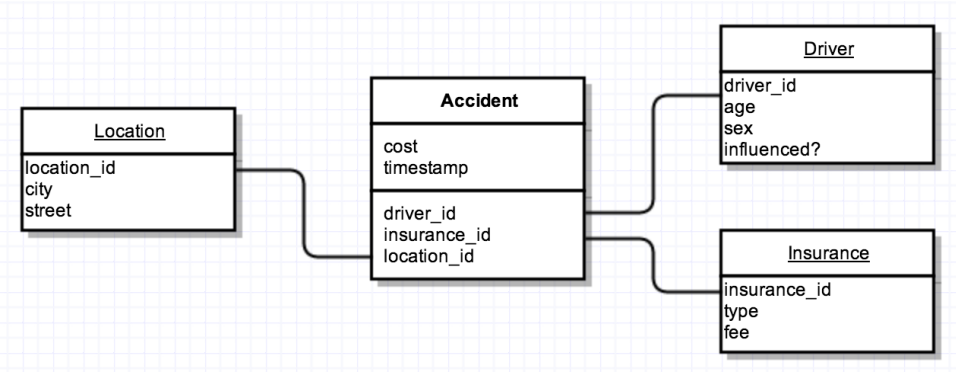
Going up the concept hierarchy, broadening the selection.

Drill-down

Going down the concept hierarchy, narrowing the selection.

2 (alternative a)

a) Make a star or snowflake schema for this case description.



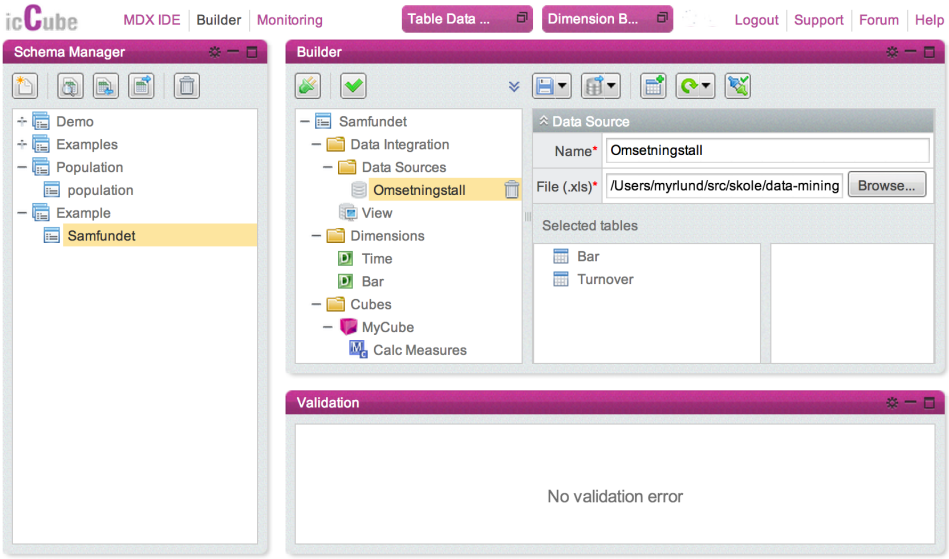
b) Define two different concept hierarchies (freely chosen dimensions).

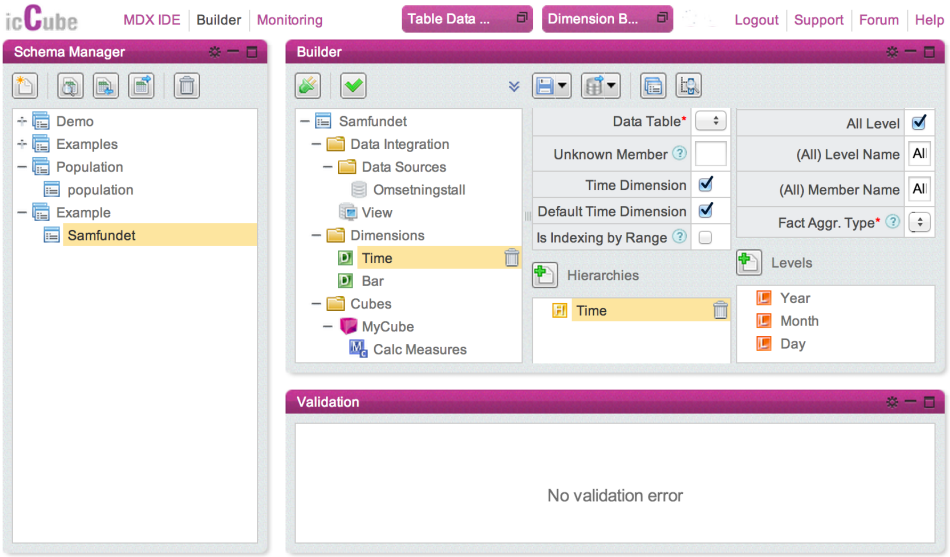
1. City → Street
2. Year → Month → Day (assuming it is extracted as from fact table)

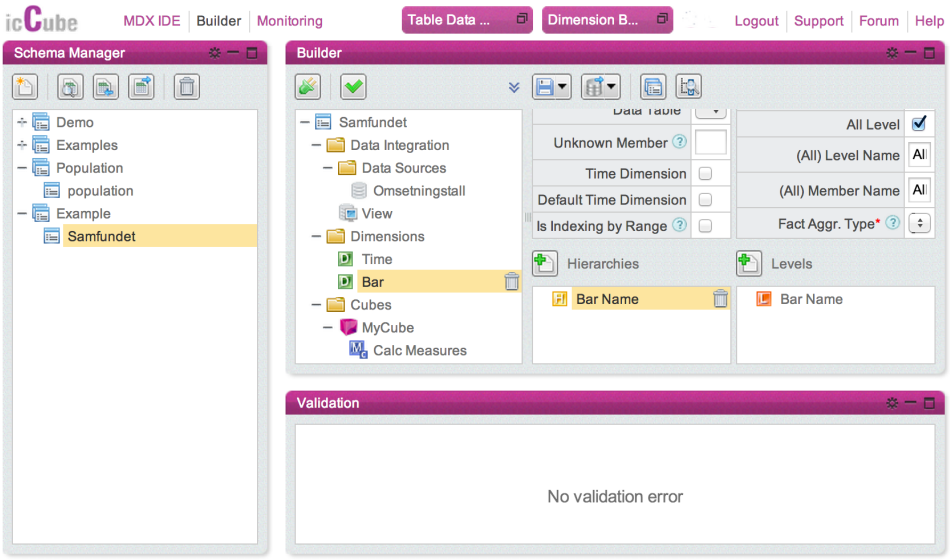
3 icCube

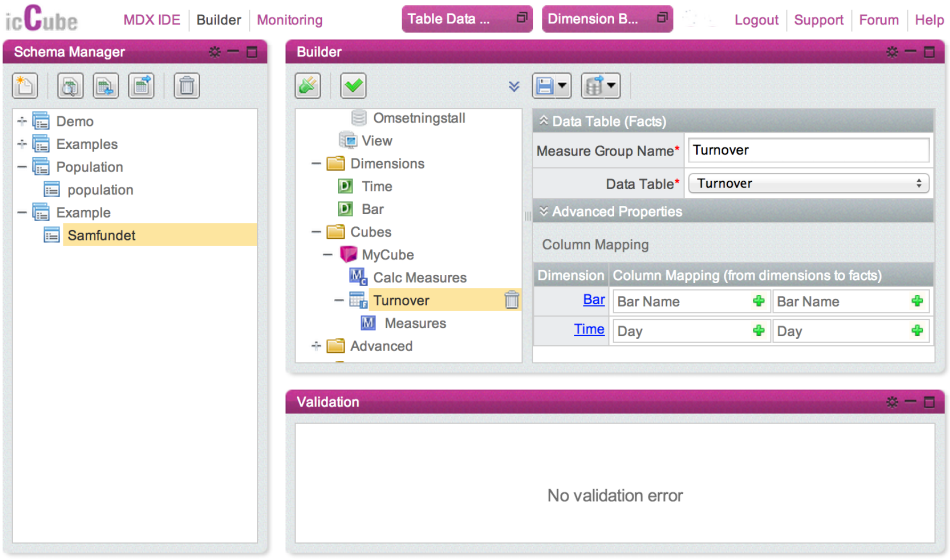
I created a highly fictuous data set with the turnovers in the bars of Samfundet through January 2014. It consists of two dimensions: time and location, while the fact table contains the turnovers.

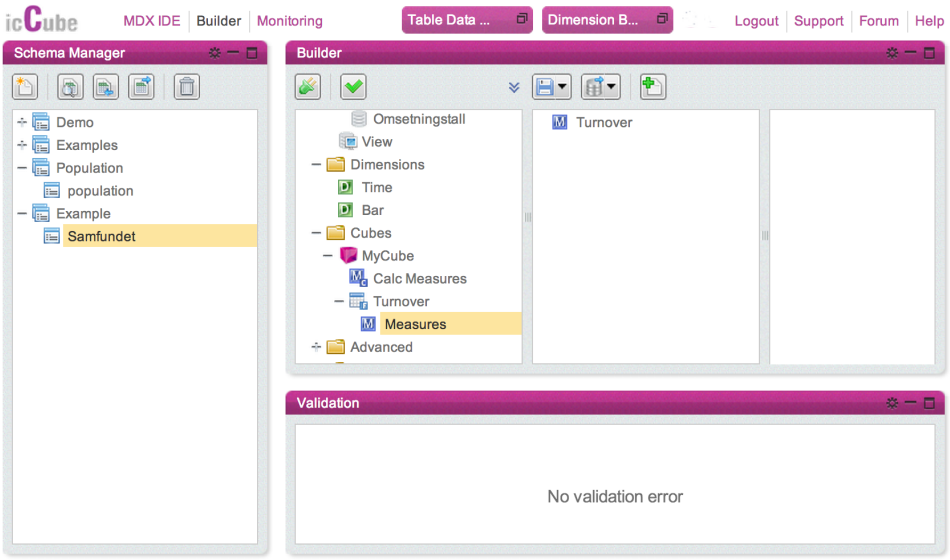
a) Setup in builder



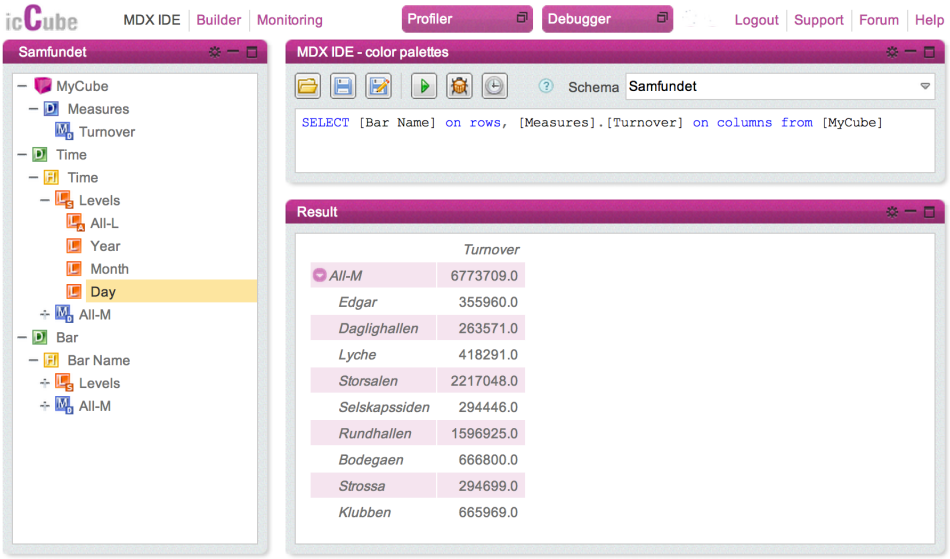


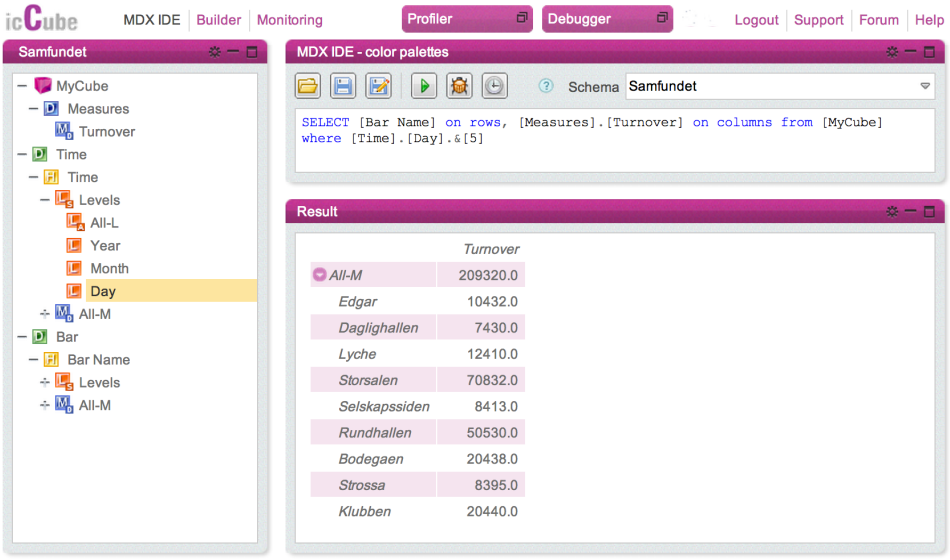

icCube
MDX IDE
Builder
Monitoring
Table Data ...
Dimension B...
Logout
Support
Forum
Help
Schema Manager
Demo
Examples
Population
population
Example
Samfundet
Builder
Samfundet
Data Integration
Data Sources
Omsetningstall
View
Dimensions
Time
Bar
Cubes
MyCube
Calc Measures
Unknown Member
Time Dimension
Default Time Dimension
Is Indexing by Range
All Level
(All) Level Name
(All) Member Name
Fact Aggr. Type*
Hierarchies
Levels
Bar Name
Bar Name
Validation
No validation error


icCube
MDX IDE
Builder
Monitoring
Table Data ...
Dimension B...
Logout
Support
Forum
Help
Schema Manager
Demo
Examples
Population
population
Example
Samfundet
Builder
Omsetningstall
View
Dimensions
Time
Bar
Cubes
MyCube
Calc Measures
Turnover
Measures
Advanced
Data Table (Facts)
Measure Group Name*
Turnover
Data Table*
Turnover
Advanced Properties
Column Mapping
Dimension
Column Mapping (from dimensions to facts)
Bar
Bar Name
Bar Name
Time
Day
Day
Validation
No validation error


icCube
MDX IDE
Builder
Monitoring
Table Data ...
Dimension B...
Logout
Support
Forum
Help
Schema Manager
Demo
Examples
Population
population
Example
Samfundet
Builder
Omsetningstall
View
Dimensions
Time
Bar
Cubes
MyCube
Calc Measures
Turnover
Measures
Advanced
Turnover
Validation
No validation error

b) Queries


icCube
MDX IDE
Builder
Monitoring
Profiler
Debugger
Logout
Support
Forum
Help
Samfundet
MyCube
Measures
Turnover
Time
Time
Levels
All-L
Year
Month
Day
All-M
Bar
Bar Name
Levels
All-M
MDX IDE - color palettes
Schema
Samfundet
SELECT [Bar Name] on rows, [Measures].[Turnover] on columns from [MyCube]
Result
Turnover
All-M 6773709.0
Edgar 355960.0
Daglighallen 263571.0
Lyche 418291.0
Storsalen 2217048.0
Selskapssiden 294446.0
Rundhallen 1596925.0
Bodegaen 666800.0
Strossa 294699.0
Klubben 665969.0


icCube
MDX IDE
Builder
Monitoring
Profiler
Debugger
Logout
Support
Forum
Help
Samfundet
MyCube
Measures
Turnover
Time
Time
Levels
All-L
Year
Month
Day
All-M
Bar
Bar Name
Levels
All-M
MDX IDE - color palettes
Schema
Samfundet
SELECT [Bar Name] on rows, [Measures].[Turnover] on columns from [MyCube]
where [Time].[Day].&[5]
Result
Turnover
All-M 209320.0
Edgar 10432.0
Daglighallen 7430.0
Lyche 12410.0
Storsalen 70832.0
Selskapssiden 8413.0
Rundhallen 50530.0
Bodegaen 20438.0
Strossa 8395.0
Klubben 20440.0