TDT4300 Assignment 1

Arve Nygård Tobias Linkjendal

07.02.2014

Exercise 1

1. a) Explain the concepts of OLTP and OLAP

OLTP, **O**nline **T**ransaction **P**rocessing, is the process of many short online transactions like `INSERT`, `DELETE`, and `UPDATE`. It is important that the query processing is fast and that the data integrity is maintained in multi-access environments. The systems effectiveness is measured by the number of transactions per second. OLAP, **O**nline **A**nalytical **P**rocessing, is handling fewer transactions than OLTP, as the queries are often very complex. The data has been analyzed and aggregated, and can tell a history. This historical data is stored in multi-dimensional schemas, like for example a star schema. The main differences between the two concepts are that OLTP focuses on operations and having detailed and current data, while OLAP focuses on the information extracted from data by analyzing it.

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

A data cube is a multi-dimensional array of values. A data cube can be used to generate cuboids consisting of a subset of the data cube's dimensions

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

- **Slice:** Picking a specific value for one of the dimensions, getting a "slice" of the cube.
- **Dice:** Picking specific values for multiple dimensions, getting a sub-cube that has been "diced".
- **Roll-up:** Viewing data in less detail, for example looking on decades instead of years. "Walking up the hierarchy"
- **Drill-down:** Viewing data in more detail, for example looking on years instead of decades. "Walking down the hierarchy"

Exercise 2 (alternative b)

2. a) Make a case description for the chosen dataset where you explain what facts and dimensions you find interesting.

We have used the provided "World population prospects" data set. This data set shows how the population in different countries will change with time, depending on the countries' fertility levels. It is interesting how we can see the population will change, and we can use this data to predict which countries and regions that will have huge or low population.

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

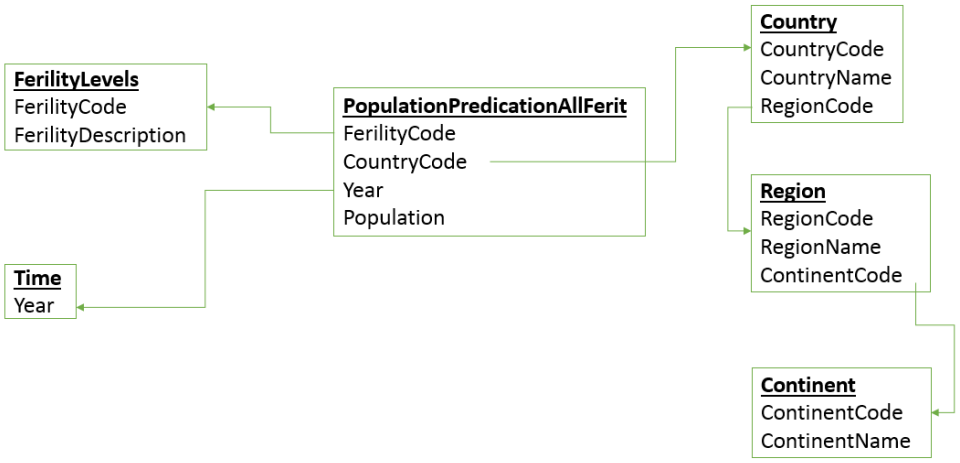


Figure 1: schema

c) Define two different concept hierarchies (freely chosen dimensions)

Geography: Continent > Region > Country **Time**: Decade > Year

3. Document your work e.g. by taking screenshots from the builder and med editor

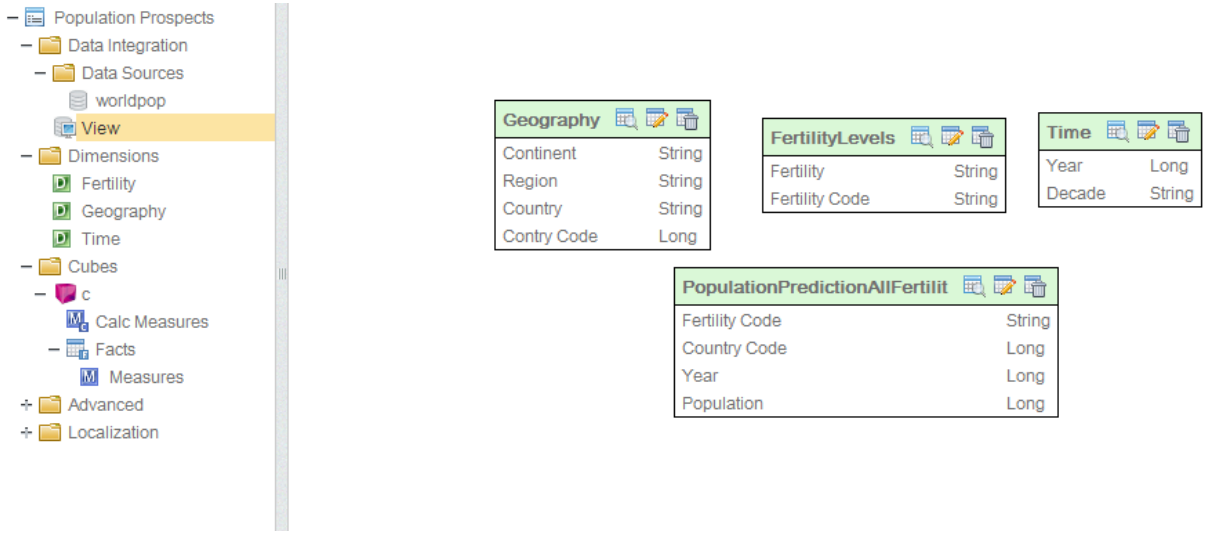


Figure 2: The tables

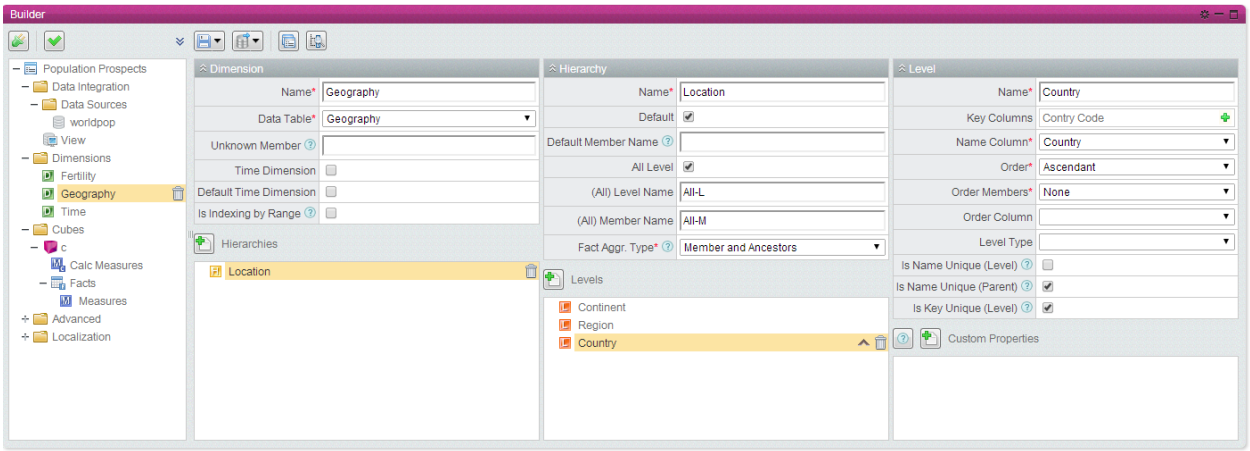


Figure 3: Dimensions and hierarchies

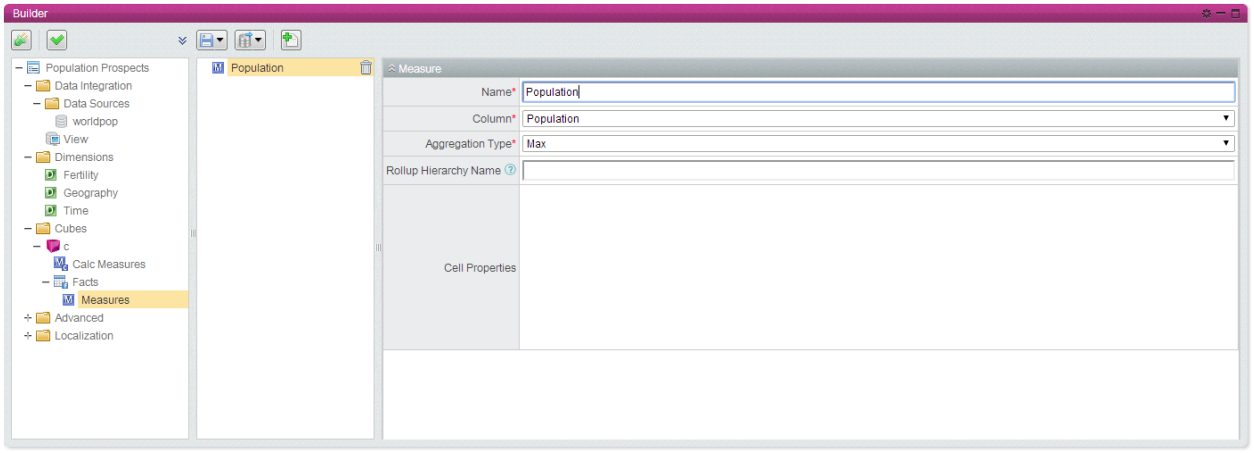
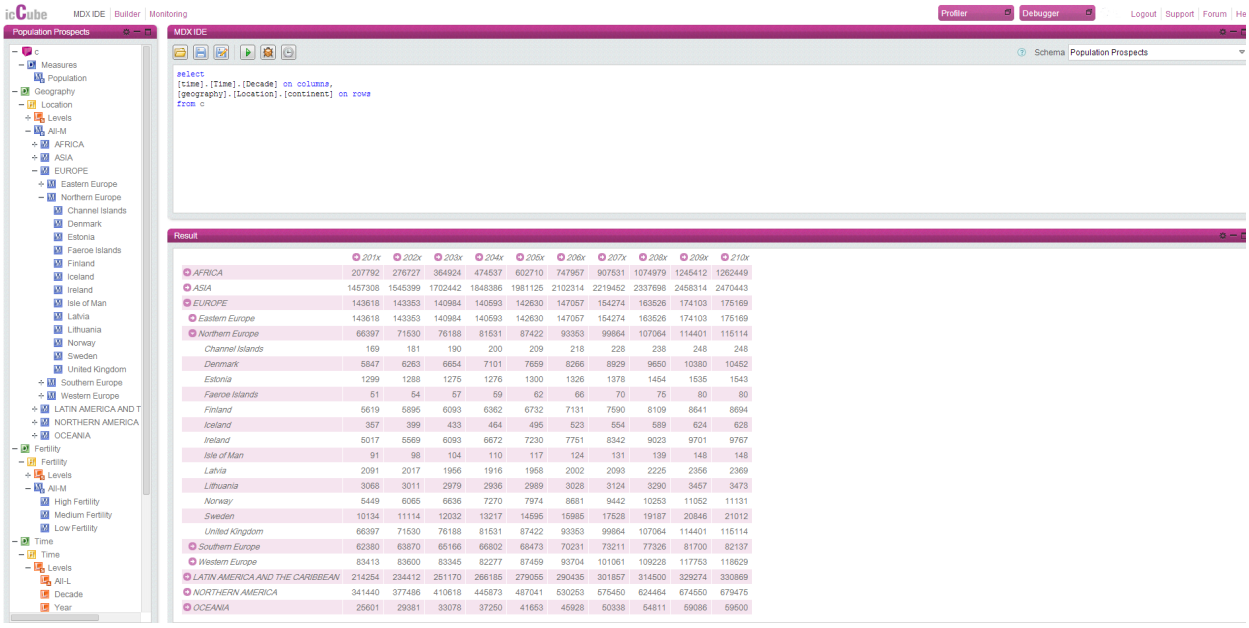


Figure 4: Facts



	201x	202x	203x	204x	205x	206x	207x	208x	209x	210x
AFRICA	207792	276727	364924	474537	602710	747957	907531	1074979	1245412	1262449
ASIA	1457308	1545399	1702442	1848386	1981125	2102314	2219452	2337698	2458314	2470443
EUROPE	143618	143353	140984	140593	142630	147057	154274	163526	174103	175169
Eastern Europe	143618	143353	140984	140593	142630	147057	154274	163526	174103	175169
Northern Europe	66397	71530	76188	81531	87422	93353	99864	107064	114401	115114
Channel Islands	169	181	190	200	209	218	228	238	248	248
Denmark	5847	6263	6654	7101	7659	8266	8929	9650	10380	10452
Estonia	1299	1288	1275	1276	1300	1326	1378	1454	1535	1543
Faeroe Islands	51	54	57	59	62	66	70	75	80	80
Finland	5619	5895	6093	6362	6732	7131	7590	8109	8641	8694
Iceland	357	399	433	464	495	523	554	589	624	628
Ireland	5017	5569	6093	6672	7230	7751	8342	9023	9701	9767
Isle of Man	91	98	104	110	117	124	131	139	148	148
Latvia	2091	2017	1956	1916	1958	2002	2093	2225	2356	2369
Lithuania	3068	3011	2979	2936	2989	3028	3124	3290	3457	3473
Norway	5449	6065	6636	7270	7974	8681	9442	10253	11052	11131
Sweden	10134	11114	12032	13217	14595	15985	17528	19187	20846	21012
United Kingdom	66397	71530	76188	81531	87422	93353	99864	107064	114401	115114
Southern Europe	62380	63870	65166	66802	68473	70231	73211	77326	81700	82137
Western Europe	83413	83600	83345	82277	87459	93704	101061	109228	117753	118629
LATIN AMERICA AND THE CARIBBEAN	214254	234412	251170	266185	279055	290435	301857	314500	329274	330869
NORTHERN AMERICA	341440	377486	410618	445873	487041	530253	575450	624464	674550	679475
OCEANIA	25601	29381	33078	37250	41653	45928	50338	54811	59086	59500

Figure 5: Query 1: Population as a function of time. Drilldown on time and geography.

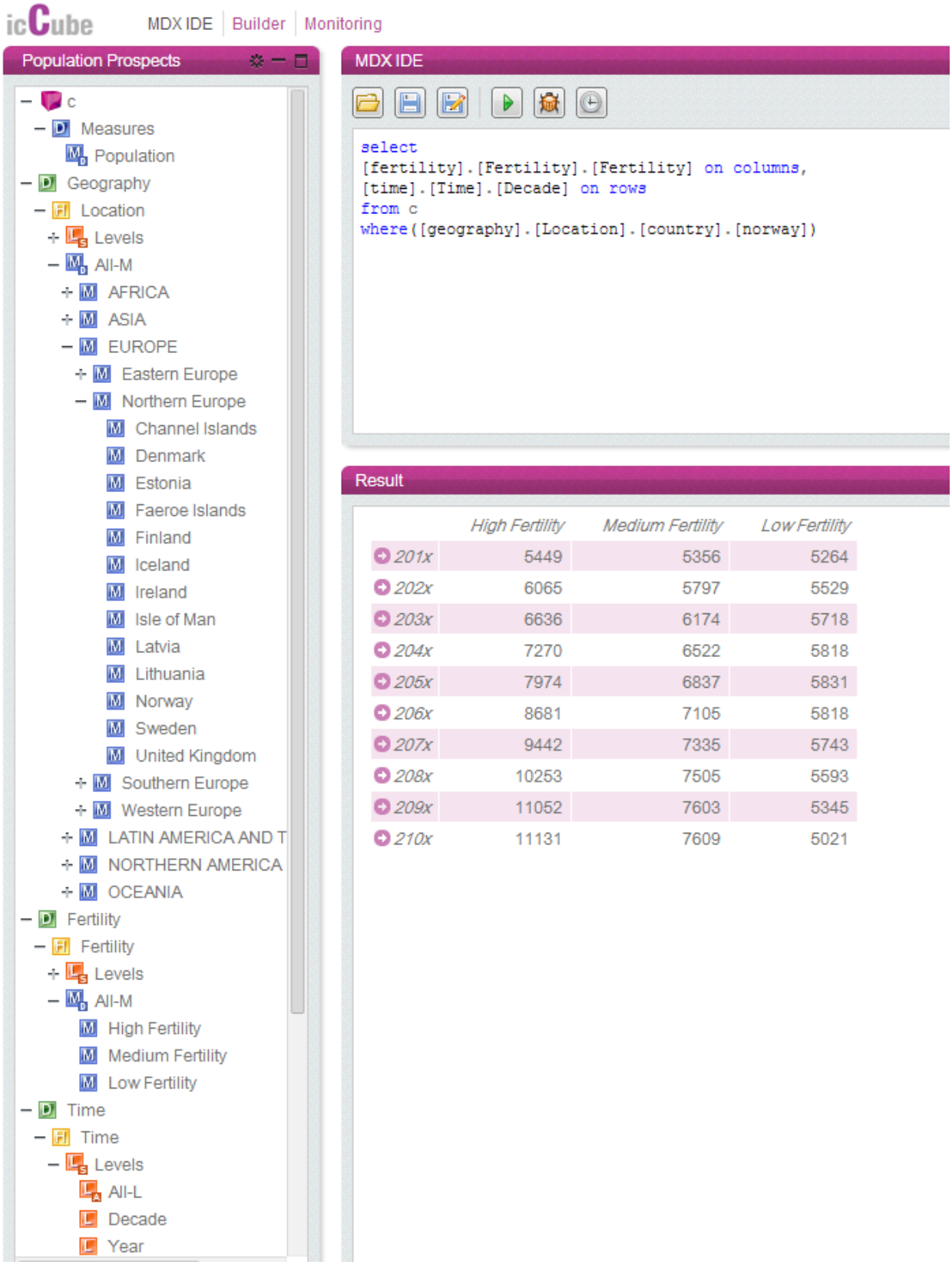


Figure 6: Query 2: Population in Norway the next 100 years, depending on our fertility.