OLAP and Data Warehousing

Lab Exercises, WS 2016/17

Multidimensional Analysis of Student Notes at FHH using OLAP Techniques

Purpose:

The purpose of this study is to
- learn how to use OLAP methods / tools to perform multidimensional analysis of data,
- get basic skills in using MS SQL Server 2014 Integration Services (SSIS) and Analysis Services (SSAS) tools for building ETL processes and designing multidimensional data cubes.

The task consists in building a simple OLAP solution for analysis of student notes at FHH. The tool will be used to investigate interesting relationships between the fact variable (note) and various dimension attributes (i.e., attributes of students, teachers, courses etc.). For instance, we will investigate whether student notes depend on such variables as teacher seniority ('Are older teachers more strict than younger teachers?'), time, Studiengang, type of course, and how this changes over time.

Input data:

Data sets:
    fhhnotes.mdb – notes at FHH

Tasks:

1. Using Integrated Services Project, design ETL process to load source data (four tables from fhhnotes.mdb) into a SQL database.
   Clean / modify / integrate data using SSIS data flow transformations.
   Useful transformations to practice:
   - Derived Column for conditional data column modification,
   - Conditional Split for conditional data subsetting,
   - Merge Joining data tables.

   Alternatively, ETL transformation can be written in SQL placed in Execute SQL Task containers in the IS package.
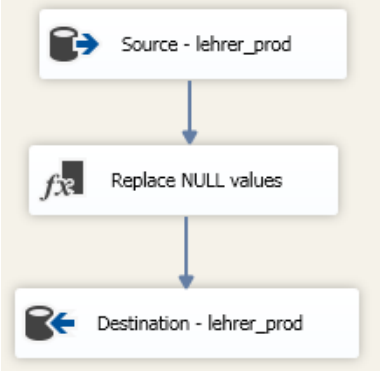
2. Using Analysis Service project, design / deploy multidimensional cube with:

   - Note as the fact variable. Following statistics should be included: number of notes and avg note. Hints 3,4.

   - All other columns available in data to be used as analysis dimensions (dimension attributes). The cube should have the following dimensions (dimension attributes), Hint 1:
     - teacher related: typ, title_code, teach_duration, teacher's sex, teacher_code, Hint2,
     - course related: lecture_group→ lecture_code hierarchy,
     - student related: student_group student_code, student's sex,
     - semester, semester_numer, semester_type (WS / SS), Hint 6,
     - probation,

o workload (of a teacher in a given semester). Hint 7.
o Where applicable, add attribute hierarchies in dimensions, e.g., semester type (WS/SS) → individual semester.

- Additional requirements:
  o Please ensure that the values in the note column are between 1 and 5 (remove illegal values). Hint 5

  o Empty values in dimension columns should be substituted with some descriptive information (e.g., a '?'). Hint 6.

3. Browse the cube, try to observe interesting relationships related to how notes depend on such classification variables as, e.g.: (consider various note statistics such as the average, number of values, range, etc.):
   - teacher senioriy (teach_duration),
   - course type,
   - job title of teacher,
   - sex of teacher,
   - probation,
   - individual teacher number,
   - time,
   - workload of teacher.

Create a report summarizing your most interesting findings.

Hints

No	How do I…	Hint
1	Add cube dimensions based on columns from the fact table	Using Data Source View Designer, create a separate table to store this column from the fact table (built as a Named Query) and link this to the fact table. The Cube Designer will use this table as the basis for another cube dimension. Alternatively, this step can be done at the ETL stage (in SSIS). See Hint 8.
2	Create a drill down hierarchy	Hierarchies of dimension attributes are built using Dimension Designer – hierarchy pane (Cube Designer → select a dimension → edit dimension to start the Dimension Designer)
3	Add measures (statistics) to be stored in a cube	Use Cube Designer – measure pane. Use the measure's AggregateFunction property to specify on what statistics (aggregate function) the measure should be based.
4	Add calculated member to a cube (to allow for e.g. AVG statistic in reports)	Add Calculated members to the cube design (Calculations tab in the Cube Designer) to enable users to analyze e.g. average note (or some other nonadditive members) based on measures stored in the cube
5	Subset data (remove unwanted rows)	Use SSIS SQL task to subset data or Conditional Split data transformation tool to achieve this.
6	Change the way values	Source View designer – add a Named Calculation to a table, e.g.,

| | of dimension are displayed (e.g., in order to change a blank into a '?') | ```
CASE
    WHEN lecture_group IS NULL THEN
        '?'
    ELSE
        CAST(lecture_group AS nvarchar)
END
```<br><br>Alternatively, replace NULL values in the ETL step using Derived Column data transformation tool:<br><br><br><br>with the following as in the following example:<br><br>| Derived Column Name | Derived Column | Expression |<br>|---|---|---|<br>| titel_code | Replace 'titel_code' | ISNULL(titel_code) ? "0" : titel_code |<br><br>Alternatively, change the NULL values in the DB table, using Execute SQL Task, e.g.,<br><br>```\nUPDATE table_name\n  SET title_code = '?'\n  WHERE title_code IS NULL\n``` |
7	Add the workload attribute	SSIS: using SQL task compute number of notes (or different courses) per teacher/semester and store this in a separate table. Later define relationship between this dimension table and the fact table using teacher/semester composite key. While adding this dimension to the cube, notice that this column has large number of values. So it is convenient to change this dimension's attributes DiscretizationMethod and DiscretizationBucketCount, so that values shown in the user interface are grouped.
8	Create new dimension table based on non-FK column(s) from the fact table	For a non-FK column in the fact table, create a new dimension table (the non-FK column in the fact table will become a foreign key to the newly created dimension). E.g., semester non-FK column in the fact table notes: ```\nSELECT DISTINCT semester (and other columns related\n to semester)\nINTO semester\nFROM notes\n```
9	Resolve the cube deployment failure due to insufficient to access	Security issue: if deployment of the cube fails due to insufficient rights of Analysis Services attempting to access the ETL database, then you need to grant the LOCALHOST\MSSQLServerMSASUser

	the ETL database	the db_datareader role for the ETL database. To do this, follow this procedure: 1. In Management Studio connect to the SQL Server engine 2. Select Security – Logins – New Login 3. Login Name – Search 4. Object Types – select Groups 5. Advanced – Find Now 6. Select SQLServerMSASUser login 7. In Login Properties for this login, select User Mapping 8. Select the ETL database (check the map box) and select the db_datareader role