# Module 2 Assignment

The assignment in module 2 involves concepts from modules 2 and 3 of course 2. As preparation for this assignment, you should review the mini case studies and associated practice problems and graded assignment in module 3 of course 2.

You should read the data warehouse design background and associated documents in module 2 before starting on the assignment. The case study for the module 2 assignment contains a primary data source and three secondary data sources with sample data for each data source. Appendices in the background document provide details about each data source. You should also see the statement of business needs for the data warehouse. Using the data sources and business needs, you should specify a dimensional model with dimensions, measures, and grain, create a schema design for the data warehouse that integrates the data sources, identify summarizability problems in the design, and populate data warehouse tables from sample rows in the data sources.

## Problems

Your design should support analysis of leads, jobs, subjobs, shipments, and invoices as well as financial summaries. The complete data warehouse design is larger than designs in the mini case studies. You may feel a little overwhelmed. To reduce your work, you may limit your data warehouse design to four data cubes and associated star or snowflake schemas. If you have more time, I encourage you to complete the entire design beyond the four items (cubes and associated table designs).

1. You should identify dimensions, map dimensions to data sources, and specify dimension hierarchies in each data cube. For each dimension, you should identify its data sources and

attributes in each data source. For hierarchical dimensions, you should indicate the levels from broad to narrow. Here is a template table to help structure your solution.

| Dimension | Attributes | Hierarchies | Data Sources |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

2. You should specify measures, related data sources, and measure aggregation properties. Here is a template table to help structure your solution.

| Data Source | Measures | Aggregation Properties |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

After identifying measures, you should put the dimensions from problem 1 and the measures from this problem into data cubes using this template table.

| Cube | Dimensions | Measures |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

3. In each data cube, you should identify the grain using the business needs as a guideline. You should then indicate relative storage requirements for the grain using the statistics for the data sources. Using the cardinality estimates provided, you should determine either the fact table

size or sparsity and then compute the unknown grain size variable. For example, you should compute sparsity if the fact table size is given.

To structure your answer, you should complete the following table. The unadjusted size is the product of the dimension cardinalities. The sparsity is the unadjusted size divided by the estimated fact table size. Put the details of your calculation in a spreadsheet. I encourage you to provide grain estimates for the finest grain and a somewhat coarser grain. For example, the finer grain may involve individual customers while the coarser grain may involve customer postal codes.

| Cube | Grain | Unadjusted Size | Sparsity |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

4. Transform your data cubes into table designs (schemas), either using star or snowflake patterns. You do not need to write CREATE TABLE statements. You should create schema diagrams using a drawing tool of your choice. For each table in a schema, you should define the table name, primary key, and columns. You should design a constellation schema that covers each component (star or snowflake) schema. You should also have a matrix or table that maps dimension and fact tables from each component schema to the constellation schema. You can use this template table to document your constellation schema. In summary, you should submit at least 4 component schema diagrams (star or snowflake) and complete the template table to document the constellation schema that combines the component schema diagrams.

| Schema type | Dimension tables | Fact table | Comments |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

5. Identify potential summarizability problems in your star schema and indicate preferred resolutions of the summarizability problems. For incomplete dimension-fact relationships, you should also indicate if columns in a dimension table allow null values.
6. You should populate your data warehouse tables based on the sample data for the ERP tables and other data sources. You can find the sample data in a spreadsheet a separate reading item with documents for assignment 2. You should not write SQL INSERT statements or insert data into database tables. Instead, you should create a spreadsheet with sample data for each data warehouse table. To help with the tedious mapping to populate data warehouse tables, you can use the sample time dimension rows in Appendix A.

### *Solution Quality*

Quality is rather subjective in data warehouse designs, but some elements are less subjective. I suggest that you address these quality items in the appropriate part of your solution.

- Schema pattern: Your design should have a constellation schema combining multiple star or snowflake schemas.
- Fact table selection: You should study fact table selection in the solution for the practice mini cases for inspiration. Typically, the fact table combines a two level design in a source schema into a single fact table. For example, an order heading and order detail are usually combined into a fact table recording the order details with dimension relationships to capture the order heading. However, in some data warehouse designs both levels of detail are needed. You should carefully consider if the data warehouse should flatten two level designs in the source databases.

- Missing data in populated tables: You should ensure that your populated tables include all data from the ERP database and lead file. The best check on your schema design is to map sample rows from the data sources to the data warehouses.
- Simplicity: Typically, a data warehouse schema design simplifies the schemas of the underlying data sources. You should consider the relevance of each source column for usage in a data warehouse. Simplification can involve combining some elements of data sources in decisions about dimensions and fact tables. Typically, relationships among fact tables are not needed in a data warehouse design even if the source schema shows relationships.

### *Grading*

The assessment method for this assignment is peer review. Each problem has an equal grade.

### *Submission*

You need to submit 6 documents to the peer review for module 2 assignment. Each document contains a full solution for the problem. You should neatly format your documents so that it can be easily graded. Please write the problem number at the top of the page.

## Appendix A: Time Dimension Table

Since the date dimension table is tedious to populate, you can use the following table. Note that this table only contains dates used in data sources so it is not a completely populated table. You should reference rows in the date dimension table for dates in fact tables. You can change the primary key if you want. The primary key in this table is a concatenation of the year, month, and day for ease of mapping.

| Time_Id | Year | Quarter | Month | Day | Week |
|---|---|---|---|---|---|
| 20130608 | 2013 | 2 | 6 | 8 | 23 |
| 20131205 | 2013 | 4 | 12 | 5 | 49 |
| 20131220 | 2013 | 4 | 12 | 20 | 51 |
| 20140101 | 2014 | 1 | 1 | 1 | 1 |
| 20140103 | 2014 | 1 | 1 | 3 | 1 |
| 20140105 | 2014 | 1 | 1 | 5 | 2 |
| 20140110 | 2014 | 1 | 1 | 10 | 2 |
| 20140115 | 2014 | 1 | 1 | 15 | 3 |
| 20140117 | 2014 | 1 | 1 | 17 | 3 |
| 20140124 | 2014 | 1 | 1 | 24 | 4 |
| 20140125 | 2014 | 1 | 1 | 25 | 4 |
| 20140131 | 2014 | 1 | 1 | 31 | 5 |
| 20140201 | 2014 | 1 | 2 | 1 | 5 |
| 20140202 | 2014 | 1 | 2 | 2 | 6 |
| 20140203 | 2014 | 1 | 2 | 3 | 6 |
| 20140215 | 2014 | 1 | 2 | 15 | 7 |
| 20140224 | 2014 | 1 | 2 | 24 | 9 |
| 20140228 | 2014 | 1 | 2 | 28 | 9 |
| 20140301 | 2014 | 1 | 3 | 1 | 9 |
| 20140314 | 2014 | 1 | 3 | 14 | 11 |
| 20140315 | 2014 | 1 | 3 | 15 | 11 |
| 20140331 | 2014 | 1 | 3 | 31 | 14 |
| 20140401 | 2014 | 2 | 4 | 1 | 14 |
| 20140430 | 2014 | 2 | 4 | 30 | 18 |
| 20140501 | 2014 | 2 | 5 | 1 | 18 |
| 20140517 | 2014 | 2 | 5 | 17 | 20 |
| 20140531 | 2014 | 2 | 5 | 31 | 22 |
| 20140601 | 2014 | 2 | 6 | 1 | 23 |
| 20140603 | 2014 | 2 | 6 | 3 | 23 |
| 20140615 | 2014 | 2 | 6 | 15 | 25 |
| 20140630 | 2014 | 2 | 6 | 30 | 27 |

| | | | | | |
|---|---|---|---|---|---|
| 20140701 | 2014 | 3 | 7 | 1 | 27 |
| 20140707 | 2014 | 3 | 7 | 7 | 28 |
| 20140731 | 2014 | 3 | 7 | 31 | 31 |
| 20140801 | 2014 | 3 | 8 | 1 | 31 |
| 20140810 | 2014 | 3 | 8 | 10 | 33 |
| 20140831 | 2014 | 3 | 8 | 31 | 36 |
| 20140901 | 2014 | 3 | 9 | 1 | 36 |
| 20140905 | 2014 | 3 | 9 | 5 | 36 |
| 20140915 | 2014 | 3 | 9 | 15 | 38 |
| 20140930 | 2014 | 3 | 9 | 30 | 40 |
| 20141001 | 2014 | 4 | 10 | 1 | 40 |
| 20141005 | 2014 | 4 | 10 | 5 | 41 |
| 20141031 | 2014 | 4 | 10 | 31 | 44 |
| 20141101 | 2014 | 4 | 11 | 1 | 44 |
| 20141120 | 2014 | 4 | 11 | 20 | 47 |
| 20141130 | 2014 | 4 | 11 | 30 | 49 |
| 20141201 | 2014 | 4 | 12 | 1 | 49 |
| 20141208 | 2014 | 4 | 12 | 8 | 50 |
| 20141215 | 2014 | 4 | 12 | 15 | 51 |
| 20141231 | 2014 | 4 | 12 | 31 | 53 |
| 20150103 | 2015 | 1 | 1 | 3 | 1 |
| 20150110 | 2015 | 1 | 1 | 10 | 2 |
| 20150125 | 2015 | 1 | 1 | 25 | 5 |
| 20150126 | 2015 | 1 | 1 | 26 | 5 |
| 20150202 | 2015 | 1 | 2 | 2 | 6 |
| 20150203 | 2015 | 1 | 2 | 3 | 6 |
| 20150213 | 2015 | 1 | 2 | 13 | 7 |
| 20150214 | 2015 | 1 | 2 | 14 | 7 |
| 20150215 | 2015 | 1 | 2 | 15 | 8 |
| 20150215 | 2015 | 1 | 2 | 15 | 8 |
| 20150224 | 2015 | 1 | 2 | 24 | 9 |
| 20150228 | 2015 | 1 | 2 | 28 | 9 |
| 20150303 | 2015 | 1 | 3 | 3 | 10 |
| 20150304 | 2015 | 1 | 3 | 4 | 10 |
| 20150315 | 2015 | 1 | 3 | 15 | 12 |
| 20150407 | 2015 | 2 | 4 | 7 | 15 |
| 20150410 | 2015 | 2 | 4 | 10 | 15 |
| 20150410 | 2015 | 2 | 4 | 10 | 15 |
| 20150411 | 2015 | 2 | 4 | 11 | 15 |
| 20150417 | 2015 | 2 | 4 | 17 | 16 |
| 20150501 | 2015 | 2 | 5 | 1 | 18 |
| 20150505 | 2015 | 2 | 5 | 5 | 19 |
| 20150510 | 2015 | 2 | 5 | 10 | 20 |
| 20150515 | 2015 | 2 | 5 | 15 | 20 |
| 20150529 | 2015 | 2 | 5 | 29 | 22 |
| 20150714 | 2015 | 3 | 7 | 14 | 29 |

| | | | | | |
|---|---|---|---|---|---|
| 20150820 | 2015 | 3 | 8 | 20 | 34 |
| 20150901 | 2015 | 3 | 9 | 1 | 36 |
| 20151017 | 2015 | 4 | 10 | 17 | 42 |
| 20151030 | 2015 | 4 | 10 | 30 | 44 |