

Dimension Tables

by Sudheer Sharma - Tuesday, November 30, 2010

<http://dwhnotes.com/archives/1552>

What do you think of this post?

[Awesome \(0\)](#) [Interesting \(0\)](#) [Useful \(1\)](#)

Dimension Tables

Dimension tables are used to describe, filter and aggregate the data that is collected for a specific activity or event. Practically they can be considered as the by words when defining the information content of a report. For example number of encounters by month by provider. Dimensions have relatively few rows compared to the fact tables, but have many descriptive textual attributes. Many times they contain hierarchies, for example location, province, country and are highly denormalized, storing even multiple hierarchies in a single table. A dimension has a primary key, that is usually a meaningless, unique integer, called a surrogate key. These keys are handled by the ETL process and are used to join dimensions to fact tables. In addition, a dimension has one or more fields that serve as the natural key of the table and are based on fields that come from the source system. When we track changes in a dimension there is a one-to-many relationship between a natural key and a surrogate key. Finally, a dimension contains many (hopefully) descriptive attributes.

Delivering Dimension Tables

At this step the dimension tables are loaded with the new and changed data. Dimension tables consist of a primary key (which is a meaningless integer key – a surrogate key), which is used to join to fact tables, a natural key which is based on fields of the source system, that doesn't change over time or across systems (e.g. for a patient dimension we use the patient identifier for natural key, not a patient id) and descriptive attributes.

We also need to handle changes to dimension attributes. There are three approaches:

1. Overwrite the existing dimension record (type 1 slowly changing dimension)

In this case we use an update else insert (merging) functionality.

2. A new dimension record is created with a new surrogate key, but the same natural key (type 2 slowly changing dimensions).

If a dimension contains attributes that are handled as above, we'll add two fields: a row effective datetime (exact datetime of change) and a row end datetime (exact datetime of next change). This way, we know the interval a record was valid.

3. Two columns are created for each attribute that we want to track changes. The first column contains the most recent value and the second column contains the old value.

For the moment Type 1 is used.

Other Issues:

- Generating surrogate keys
- Date Dimension
- Dimensional Roles

What do you think of this post?

[Awesome \(0\)](#) [Interesting \(0\)](#) [Useful \(1\)](#)

PDF generated by Kalin's PDF Creation Station