Instructor Information:

- Laura L. Reeves, co-author of *The Data Warehouse Lifecycle Toolkit* has over 22 years of experience in end to end data warehouse development focused on developing comprehensive project plans, collecting business requirements, developing Business Dimensional Models, database schemas (both star and snowflake designs), and development of enterprise data warehouse strategies. As StarSoft Solutions co-founder, Laura has implemented data warehouses for many business functions for private and public industry.

Laura L. Reeves  
1163 East Ogden Ave.  
Suite 705-369  
Naperville, IL  60563  

Phone: 630.369.2787  
FAX: 630.369.4740  
laura@starsoftinc.com
Topics

• Introducing the technique
  ➜ The Business Dimensional Model (BDM)
• BDM fundamentals
• Translating the BDM
• More BDM notation
• Putting it all together
Disconnect between Business and IT

- **Business Perspective**
  - IT does not deliver what we need
  - Build solutions ourselves

- **IT Perspective**
  - Business can’t tell us what they really want
  - Build it for them and expect them to use it

- **May get lucky and meet short term needs, but usually limits long term success**
Building a Bridge

- Gather broad requirements
- Partner to create the data model
- Work in business terms
- Keep technology out of the discussion
The Purpose of Dimensional Models

- **End user access**
  - Ease of use
  - Query performance
- **Delivered through a variety of technologies**
  - Relational database
    - Star or snowflake schemas
  - Multi-dimensional database
    - Cube
  - Other proprietary databases
    - Columnar
    - DW appliance

- Dimensional modeling for end user access has been effective for more than two decades. These models are being supported by data access and database software products. Great improvements are being made regularly for improved performance, increased scalability and increased functionality.
The Modeling Technique

• The Business Dimensional Model™
• Reflect data model visually and in business terms
• Ensures support for various types of reporting and analysis
• Independent of technology
• Ideal training tool for use later
• Manageable approach to complex environments

• The Business Dimensional Modeling technique was developed by StarSoft Solutions, Inc. Although this is a trademarked approach, there are no licensing fees or permissions needed to use it. All we ask is that if you choose to utilize this approach that you include a trademark reference (Business Dimensional Lifecycle is a trademark of StarSoft Solutions, Inc.).

• Overall, use any technique that is effective for you and your organization. Just make sure that all of the design objectives are covered.
Basics of Dimensions

- Dimensions are major business categories or groupings to describe business data
- Dimensions contain:
  - Descriptive elements
  - Reference data
  - Attributes, hierarchies, drill paths
- Dimensions are used for:
  - Selection criteria
  - Report labels
  - Pick list via interface

- Dimensions are what provide the users with the flexibility and variety of ways to slice, group and organize the data. You should try to provide as many attributes as you can in the dimensions. If the data source you are extracting from has only minimal descriptive data, look for other reference sources to use.
• The Business Dimensional Modeling notation reflects the lowest possible grain with a shaded box, visually located at the bottom of the page.

• Relationships between attributes are reflected with an arrow. Do not be concerned that there are no ‘crows feet’. The business folks will not notice, but the arrow head on line represents a ‘many’ relationship. Most dimensional relationships are one to many.

• Each hierarchy is represented visually moving upwards.

• Other attributes fill in any remaining open space. Try to keep the overall diagram to be balanced.
### Complete Dimension Documentation

<table>
<thead>
<tr>
<th>ATTRIBUTE NAME</th>
<th>BRIEF DESCRIPTION</th>
<th>SAMPLE VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>Calendar day information was reported on.</td>
<td>9/30/2003, 9/30/2003</td>
</tr>
<tr>
<td>Holiday</td>
<td>This is a future attribute to indicate that this date is considered a U.S. business day off.</td>
<td>Memorial Day, President's Day</td>
</tr>
<tr>
<td>Fiscal Week</td>
<td>The week of the fiscal year that the date rolls up to and is used for reporting purposes.</td>
<td>FY 2003 Week 16</td>
</tr>
<tr>
<td>Fiscal Month</td>
<td>The month of the fiscal year that this date rolls up to and is used for reporting purposes.</td>
<td>FY 2003-10, FY 2003-11</td>
</tr>
<tr>
<td>Fiscal Quarter</td>
<td>The quarter of the fiscal year that this date rolls up to and is used for reporting purposes.</td>
<td>FY 2003-Q1, FY 2003-Q2</td>
</tr>
<tr>
<td>Fiscal Year</td>
<td>The year that this date rolls up to based upon the company's fiscal calendar.</td>
<td>FY 2003</td>
</tr>
<tr>
<td>Calendar Month</td>
<td>The calendar month used for reporting purposes.</td>
<td>March 2003, October 2003</td>
</tr>
<tr>
<td>Calendar Quarter</td>
<td>The calendar quarter used for reporting purposes.</td>
<td>CY 2003-Q1, CY 2003-Q2</td>
</tr>
<tr>
<td>Calendar Year</td>
<td>The calendar year used for reporting purposes.</td>
<td>CY 2003</td>
</tr>
<tr>
<td>Week</td>
<td>The calendar week ending Saturday and is used for reporting purposes.</td>
<td>Week Ending 11/21/1998</td>
</tr>
<tr>
<td>Day of Week</td>
<td>Attribute to indicate the calendar day of week.</td>
<td>Wednesday, Saturday</td>
</tr>
</tbody>
</table>
Basics of Facts

- Facts are the basic business events that are measured and monitored
- Facts are typically:
  - ‘Amounts’ and ‘Counts’
  - Defined by more than one dimension
  - Numeric (frequently, but not always)
- Facts are used as:
  - Body of reports
  - Part of calculations

The facts are the heart of the entire system. This is where it all comes together. These are the fundamental measurements of the business. You must ensure that you determine the lowest level of detail that is possible for these facts. Once the data is loaded, you can always roll up the facts to many different levels. However, it is much more difficult to break the numbers down at query time.
• We see here that the name of the facts included in this group is in the center of the diagram. Each dimension that applies to this fact group is shown in the circles surrounding the center. The specific name of the dimension and the level of detail is included for each dimension.
Complete Fact Group Documentation

<table>
<thead>
<tr>
<th>FACT NAME</th>
<th>FACT DEFINITION</th>
<th>AGGREGATION RULE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dollars</td>
<td>The dollar amount for the Units Sold.</td>
<td>Sum</td>
</tr>
<tr>
<td>Units</td>
<td>The number of consumer units sold.</td>
<td>Sum</td>
</tr>
<tr>
<td>Price</td>
<td>The actual price charged for this sale.</td>
<td>Average</td>
</tr>
<tr>
<td>Cost</td>
<td>The standard price factor for this item sold at this store.</td>
<td>Average</td>
</tr>
</tbody>
</table>
**Dimensional Model** - A data model organized for the purpose of user understandability and high performance. In a relational database, a dimensional model is a star join schema characterized by a central fact table with a multi-part key. The components of this key are joined to a set of dimension tables, each defined by their own primary keys. In Multi-dimensional database, a dimensional model is the database cube.

**Star Schema** - An organization of tables in a relational database with a single central “fact” table possessing a multi-part key, together with a set of surrounding “dimension” tables each possessing their own primary keys. The fact table usually represents a set of measurements or “facts” supplied by an operational computer system. The most useful facts are numeric and additive. The dimensional tables are usually populated with text or text-like attributes describing specific entities like products or customers.
A BDM Case Study
Now, we will walk through a sample Business Dimensional Model for a telecommunications outsourcing business. Then we will look at the resulting logical table structures.
The Tel-Sell Call Center Dimension is represent by this slide and has a very simple hierarchy. This shows how the individual call centers are described and their organizational structure.
• The Employee Dimension is represented by this slide and again it was very straightforward. Since employees move around the organization regularly, this needed to be separate from the call center information.
• The Tel-Sell Client Dimension is represented by this slide and begins to bring in complexity. This represents the clients and the individual programs that are run, ranging from 30 days to more than 1 year in length.
• The Tel-Sell Call Status Dimension is actually a very complex dimension although the hierarchy is quite simple. This dimension represents the possible results from an individual call (sold a product, got a busy signal etc.) The challenge comes in with identifying the specific elements within each attribute level. A lot of discussion was centered here. This dimension actually contains a series of hierarchies that have different levels of detail. Each of these hierarchies would end at a different level. Combined with the multiple ways that the detailed call dispositions are combined, a simple hierarchy was agreed upon. This allows endless possibilities for standard and custom groupings of the call dispositions.
• The Activity Dimension represents how employees spend their time. Since employee costs are the highest percentage of overall costs, it is critical to understand how individual customer service representatives spend their time. For example, on the phone (in this case this is the desired behavior!), in training or on break.
• Facts about calls are described by all of the dimensions except Activities.
• Facts about hours worked are described by all of the dimensions except call status.
Using the Model

- What are the total number of hours worked to date for people who were hired this year?
- What are the total number of hours worked by fiscal year and call center region?
- What are the total calling minutes by call disposition grouped by category?
- What are the average number of calls for client “Our Biggest Customer” by fiscal week?

Does the BDM support the following questions?

- What are the total number of hours worked to date for people who were hired this year?

- What are the total number of hours worked by fiscal year and call center region?

- What are the total calling minutes by call disposition grouped by category?

- What are the average number of calls for client “Our Biggest Customer” by fiscal week?
Translating the BDM
Implementation of BDM

- Exploit the DB and BI tools
- Technical design driven by tools
  - BI Tool is the user’s only perspective of the DW
  - Database technology helps determine data structure design
- Every tool has its own unique requirements, recommendations and/or preferences

Don’t lose any sleep over your desire to develop the independent data model – I can assure you that if you change technologies in the future, the dimensions will be the least of your concerns. Yes, there will be some effort to split apart or consolidate a dimension, but the hard work is finding and populating those attributes to begin with! Also, the effort to convert your systems created applications AND all of the user created queries will be a much bigger challenge than re-configuring the dimension table structure!
• Note that each dimension is represented as a single table.
• Again each dimension is represented by a single table. Note that the Period, Employee, Client, and Call Center dimension tables are the same. This means that these dimensions are conformed.
Too Many or Too Few Dimensions?

- Practical Guidelines for a single fact table
- Minimum 3 – 4 dimensions
  - Split dimensions that are too broad
  - Consider enterprise view, would that split dimensions apart?
  - Are there more dimensions that could be included for user benefit, but not change the number of rows in the fact table?
- Maximum 20 Dimensions
  - Are these dimensions really separate?
  - Walk through logical drill paths

Selecting dimensions is sometimes difficult up front. There you sit with the file layouts from your source systems and you are trying to get started. To get started with what dimensions you will have, think about having a big pile of elements dumped on the table. Then start dividing them up, sorting them. If you had a pile of Legos, you could sort on color, type of shape (regular shapes vs. unusual parts), size of the piece. You are doing the same thing with your data elements. You may want to sort one way – creating 3 dimensions. As you try to put other elements into these dimensions, they may not fit in very well. Perhaps you have found a new dimension. Also, after you get a group of elements that you think are separate, you may see that these are simply a hierarchy for another dimension. If so, these should be combined.

Explore what the model looks like if these dimensions are together, separate, together, separate…. One may become obvious. In some cases you may not make a final call until you have created a test database and tried to create some reports.

Another great source to determine what really belongs together are the business users. They often have clear opinions about what makes sense to them.

- One way to get their input (without asking Do you think this is one or two dimensions?), is to walk through several scenarios. If they saw a report for the Total US and ‘drilled down’ what would they expect to see? Regions. OK, now drill again. Keep this up. These drill paths are typically within a single dimension.

- You can indeed get from the lowest level of one dimension (Sales Territory) to Sales Rep (in a separate dimension). However, the interface to a user is significantly different! From a double click (most common drill down implementation) to swapping out one attribute for another attribute. Some tools allow a drill anywhere, but again the interaction is different.

- Drill down action is a different action than changing dimensions. The user can still get the all of the data, but will it make sense?
• Comprehensive and enterprise perspective drives expanded scope for models
• Models often have
  ➔ 25+ fact tables
  ➔ 15 – 30 dimensions
• Keep in mind that not all fact tables use all of the dimensions
  ➔ In general each fact group has 5-15 dimensions

• The days of developing a single star schema to get started are long gone. Your organization may already be on a 2nd, 3rd or 4th iteration of delivering decision support (maybe with dimensional data marts maybe not).
• The demands are great, the data volumes are enormous and the business wants it yesterday (and the vendors tell you that you can implement it all tomorrow)!
• These models can be quite complex. Most users have an interest in only a subset of the model. Very few people in the organization will ever see and/or use the whole thing.
• Design broadly
  ➔ Multiple user groups for same data
  ➔ Next potential data sources
  ➔ Enterprise viewpoint
    – Plan for conformed dimensions

• Implement in several increments
  ➔ Don’t tackle too much at one time
  ➔ Break into phases
    – Each should deliver business value
  ➔ Review and update the model each time

• Although it may take some work to get permission to develop a more comprehensive model up front, the work is well worth the effort.
  • Keep yourself from designing yourself into a corner!

• Instead of having your iterations be continuous re-work of the same data until you get it right, have your iterations be implementations of the next set of valuable data. This can only be done quickly if you have done enough work up front.
More BDM Notation
• Unclear groups of attributes
  ➔ Business has not defined these yet
  ➔ To come from a new operational system not designed yet
  ➔ Indicates where these pieces are likely to fit into the model in the future

• We may not have all the answers when we are working on the model. In fact, we may be about done with what we can do and there may still be large areas that are not well defined.

• If these undefined areas are part of the core purpose for your project you have a BIG problem! Revisit your scope, seek out alternate data sources, work with the business (setting expectations and get their help).

• There are cases, where there may be a lot of good stuff in the future, but the users don't know yet. As long as this is beyond the immediate scope, you can note where these would fit (to the best of your knowledge at this time). This gives you a great head start when these are defined in the future. It also increases the confidence of the users that you haven’t forgotten their ideas for the future.
**BDM Advanced Notation**

**Derived Attributes**

- When there is not a business element with an identifier at this level
- Used to tie together several hierarchies
  - Perhaps for junk dimensions

- This can be used when we need to tie together parts of a dimension, but there is not an existing element in the data to help us. This derived attribute may be stored under the covers, the users don't have to see this. In some cases, it would be valuable to create a description for this new element and make it available to the business community.
BDM Advanced Notation
Triggers for Type 2 SCD

- Used to identify those attributes that MUST result in the creation of a new version for that row in the dimension
- Helps us determine which attributes can be type 1 and which trigger type 2

- This notation is helpful to note which attributes are critical to trigger the generation of a new version (type 2 slowly changing dimension). The shadow is intended to create the impression of multiple versions.
BDM Notation Summary

- **Dimension Grain**
- **Attribute**
- **Derived Attribute**
- **Future Attribute**
- **Slowly Changing Attr**
- **Fact Group**
- **Anchor Attribute**

**M-M Relationship**

**1- M Relationship**
Putting It All Together
Keys to Success

- Dimensional modeling provides the basic foundation for end user access and analysis
- Focus on the business requirements and the business user’s perspective
- Designs should reflect clear business dimensions and facts
- Resulting table structures will likely be influenced by technology
Coming soon...

Available late spring 2009

A Manager’s Guide to Data Warehousing

Laura L. Evers

Copyright 2009 StarSoft Solutions, Inc.