Gathering Requirements for Effective Dimensional Modeling

Margy Ross
Dec 17, 2003

© 2003 Kimball Group  All rights reserved.
Kimball Group

“Definitive Source for the Dimensional Approach”

Consulting
Project assessments and strategy
Requirements analysis
Dimensional modeling and design reviews
Technical architecture

margy@ralphkimball.com
952.974.0404

Education
Public and on-site classes
Dimensional Modeling
Data Warehousing Lifecycle

www.kimballuniversity.com
Reference Materials

- Course materials adapted from...
  - *The Data Warehouse Lifecycle Toolkit*
    - R. Kimball, L. Reeves, M. Ross, W. Thornthwaite (Wiley 1998)
  - *The Data Warehouse Toolkit, 2nd Edition*
    - R. Kimball, M. Ross (Wiley 2002)
  - *Kimball University*
    - Design Tips
    - Intelligent Enterprise articles
    - www.KimballUniversity.com
Session Agenda and Goals

- Agenda:
  - Introductions
  - Lifecycle Approach
  - Dimensional Model Overview
  - Gathering Requirements for Dimensional Modeling

- Today’s Goals:
  - Learn proven approach for gathering data warehouse business requirements
  - Learn specific concepts and practical techniques
Top 10 Reasons Data Warehouses Miss the Mark

- Focus on technology and data rather than business requirements and goals.
- Fail to secure business sponsorship.
- Pursue all-or-nothing approach rather than compelling iterations.
- Spend time and energy building normalized data foundation, with nothing left for presentation.
- Concentrate on operational performance and development rather than query performance and ease of use.
Top 10 Reasons Data Warehouses Miss the Mark

- Make the user view of the data too complex.
- Build isolated data silos (marts or warehouses).
- Only load summarized data into the user presentation area.
- Presume business requirements are static.
- Fail to recognize that the data warehouse’s success is tied directly to user acceptance.
Business Dimensional Lifecycle Approach
Business Dimensional Lifecycle

- Business Requirement Definition
- Technical Architecture Design
- Product Selection & Installation
- Physical Design
- Data Staging Design & Development
- Deployment
- Maintenance and Growth
- Project Management
- Analytic Application Specification
- Analytic Application Development
- Dimensional Modeling
- Project Planning
Dimensional Modeling Concepts
Simplified Elements of Data Warehouse

Source Systems

Data Staging Area

Presentation Area

Data Access Tools

Design Goals:
- Staging throughput
- Quality / consistency

Services:
- Transform from source-to-target
- Maintain conform dimensions
- No user query support

Data Store:
- Flat files or relational tables

Design Goals:
- Dimensional Atomic AND summary data
- Business-process centric

Data Warehouse Bus:
- Conformed facts and dimensions

Services:
- Transform from source-to-target
- Maintain conform dimensions
- No user query support

Data Store:
- Flat files or relational tables

Design Goals:
- Ease-of-use
- Query performance

Ad Hoc Query & Analysis Tools

Standard Reports

Analytic Applications:
- Modeling
- Forecasting
- Scoring
- Data mining
Terminology: Dimensions

- Set of attributes (columns) related to a subject/object
  - Who, what, when, where, why, how
  - Product, Customer, Date, Patient, Vendor, Facility, ...

- Each dimension row is a unique occurrence
  - One row per product, customer, day, ...

- Dimension attributes:
  - Report labels and query constraints
  - “By” words and “where” clauses
  - Verbose descriptive attributes, in addition to codes
  - Hierarchical relationships

PRODUCT KEY

- Product Desc.
- SKU #
- Size
- Brand Desc.
- Class Desc.
Terminology: Facts

☐ Result from a business process or business event
  - Facts are usually numeric and additive

☐ Granularity/grain
  - Identifies the fact level of detail
  - One row per sale, one row per service call, one row per claim, ...
  - Atomic grain is most flexible

- DATE KEY
- PRODUCT KEY
- STORE KEY
- PROMOTION KEY
- $ Sales
- Unit Sales
Terminology: Dimensional Model or Star Schema

- Fact table per business process / event, plus relevant dimensions

- Benefits:
  - Easier to understand
  - Better performance from fewer joins
  - Extensible to handle change
## Sales Rep Performance Report
### Central Region

<table>
<thead>
<tr>
<th>District</th>
<th>Rep</th>
<th>Jan 2003 Dollars</th>
<th>Feb 2003 Dollars</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chicago District</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adams</td>
<td></td>
<td>990</td>
<td>999</td>
</tr>
<tr>
<td>Brown</td>
<td></td>
<td>900</td>
<td>999</td>
</tr>
<tr>
<td>Frederickson</td>
<td></td>
<td>990</td>
<td>999</td>
</tr>
<tr>
<td><strong>Minneapolis District</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andersen</td>
<td></td>
<td>950</td>
<td>999</td>
</tr>
<tr>
<td>Smith</td>
<td></td>
<td>950</td>
<td>999</td>
</tr>
<tr>
<td><strong>Central Region Total</strong></td>
<td></td>
<td>4,780</td>
<td>4,995</td>
</tr>
</tbody>
</table>

**“Dimensions”**
Report, row and column headings

**“Facts”**
Numeric report values
Terminology: Conformed Dimensions

- One row for each instance of a business subject or object (product, customer, etc.)
- All fact tables use same standard dimensions
  - Established via Bus Matrix, enforced in ETL
- Consistent – apples to apples across processes
Terminology: Enterprise Data Warehouse Bus Architecture

- Purchase Orders
- Store Inventory
- Store Sales

Date  Product  Store  Promo  Distr. Center  Shipper  Vendor

© 2003 Kimball Group  All rights reserved.
Terminology: Data Warehouse Bus Matrix

- Rows = Business processes
- Columns = Conformed dimensions

<table>
<thead>
<tr>
<th></th>
<th>Date</th>
<th>Product</th>
<th>Store</th>
<th>Promo</th>
<th>Dist Ctr</th>
<th>Shipper</th>
<th>Vendor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Store Sales</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Store Inventory</td>
<td>☑</td>
<td></td>
<td>☑</td>
<td>☑</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Store Deliveries</td>
<td>☑</td>
<td></td>
<td>☑</td>
<td>☑</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist Ctr Inventory</td>
<td></td>
<td></td>
<td>☑</td>
<td>☑</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist Ctr Delivery</td>
<td></td>
<td></td>
<td>☑</td>
<td>☑</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purchase Orders</td>
<td></td>
<td></td>
<td>☑</td>
<td>☑</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Creating Conformed Dimensions

- Assign surrogate key to every dimension row
  - Buffer the DW from operational system changes
  - Integrate data from multiple sources
  - Query performance advantages
  - Prerequisite for slowly changing dimension
- Combine all attributes into Master dimension table
- Use the Master to map surrogate keys to fact rows

<table>
<thead>
<tr>
<th>Product KEY</th>
<th>Marketing</th>
<th>Logistics</th>
<th>Cost Acctg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product Code</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Width</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Cost</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dimensional Modeling Process

- Develop the Data Warehouse Bus matrix
- Follow the 4-step method
  - Step 1: Identify the business process (matrix row)
- Research data sources, including direct data exploration
- Complete the 4-step method
  - Step 2: Declare the grain
  - Step 3: Identify the dimensions
  - Step 4: Identify the facts
- Diagram the dimensional model
Key Input to Dimensional Model

Business Requirements

Data Realities

Dimensional Model:
Business Process Grain Dimensions Facts
Gathering Business Requirements
Importance of Business Requirements

- Technical Architecture
- Dimensional Model
- Physical Design
- Data Staging Design
- Application Specification
- Deployment Plan
- Maintenance and Growth
- Project Scope

Business Requirements
Recommended Approach for Uncovering Requirements

- Start with business users to understand...
  - Business objectives
  - Information/analysis themes
  - Decision-making process

- Interweave data “reality” meetings with source system experts or DBAs
  - Ability to support user themes
  - Data gotchas
Techniques for Uncovering Requirements

- Interviews
  + All voices heard
  + Less interviewee time required

- Facilitated Group Sessions
  + Less elapsed time, but more participant time
  - Limited # participants
  + Brainstorming, consensus and issue resolution

- Interviews AND Facilitated Groups
Don’t presume you already know it all

- Research available resources
  - Annual report, marketing literature, web content
- Organization chart
- Previous warehousing initiatives

Identify interview team roles

- Lead interviewer, scribe and observers

Prepare interview questionnaire

- Specific to interviewee level and function
- 1-page fallback device, not script
Preparation: Select Interviewees

☐ Horizontal user representation (breadth)
  - Understand common vocabulary and data across core functions
  - Critical to Data Warehouse Bus framework

☐ Vertical user representation (depth)
  - Vision → tactics; full perspective in target area
  - Executives, managers and analysts

☐ Preliminary IT representatives
  - Source experts, DBAs, IT liaisons and IT mgmt
Preparation: Schedule/Prepare Interviewees

☐ Scheduling logistics:
  - Solicit administrative help
  - Individual sessions for execs
  - Individual or small groups for others
    ➔ Max 2 - 3 people; homogeneous; < 2 levels
  - Maximum 3 - 4 interviews per day
    ➔ One hour for individuals; 1 1/2 hour for groups
    ➔ Minimum 1/2 hour between interviews

☐ Prepare users:
  - Conduct User Kick-off Meeting - “ownership”
  - Follow-up with pre-interview letter
    ➔ Bring copies of key analyses
Interview Flow: Introduce the Interview

☐ Set the stage…
  - Review project and interview objectives
  - Introduce team players and roles
  - Confirm time

☐ Establish overall tone
  - Practice in advance
  - No technical jargon; No phones or pagers

☐ Get users to talk about what they do
  - Job responsibilities and organizational fit?
Big-picture understanding and vision

- Objectives of your organization? What are you trying to accomplish?
- How do you measure success? How do you know doing well? How often measured?
  - Identify key business processes and facts
- Key business issues? Vision for organization?
- Opportunities to better leverage information within organization? Impact on business?
  - Identify expectations and business benefits
Similar to Business Exec, but more detailed

- Objectives of your department? How measure success? Key metrics? How often?
  - Identify key business processes and facts
- How distinguish between products? Natural way categorize products? How narrow list of products?
  - Identify dimension attributes and hierarchies
- Types of routine/ad hoc analysis performed? How often? Improvements to current methods?
  - Identify data access tool requirements & application templates
- Opportunity to impact business with improved access to information? Financial impact?
Business Analyst Report / Spreadsheet Review

- Understand current analyses and opportunities for improvement
  - What data on the report is important?
  - How do they use the report today?
  - If report were dynamic, what’s different?

Remember: Designing analytic environment, not reporting system
- Resist temptation to focus on top five reports
IT Data Audit
Interview Content

☐ Feasibility of supporting requirements
  - Overview key operational source system
  - Update frequency? Availability of historical data?
  - Known data caveats or quality issues?
    ➔ Identify data availability and gotchas

☐ Dig until confident that core data exists
  - More detailed data analysis will follow for dimensional modeling
  - Beware of classic “profitability” data trap
Interview Flow: Wrap-Up

☐ Ask about project success criteria
  - What is the #1 thing the project must accomplish to be deemed successful?
    ➔ Identify measurable success criteria

☐ Review next steps
  - General disclaimer to manage expectations
  - Deliverables and next feedback opportunity

☐ Thank for participation
Interview Flow: Ground Rules

☐ Remember interview role
  - *Do LISTEN; don’t defend, sell, try to impress, etc.*

☐ Establish peer basis
  - *Use their vocabulary*
  - *Don’t dive too quickly*

☐ Strive for conversational flow

☐ Verify communication
  - *Capture terminology precisely*

☐ Maintain interview schedule flexibility
  - *Can add new folks, but avoid interview burn-out*
Post-Interview: Review Interview Results

- Informally debrief with team
  - Common themes
  - Do-ability
  - Areas requiring clarification
  - User analytical / technical sophistication

- Fill-in rough interview notes
  - Highlight key points and vocabulary
Post-Interview: Publish Deliverables

☐ Don’t overlook formal documentation
  - Validation
  - Reference material

☐ Individual interview write-ups
  - Summary, not transcript
    ➔ Business Objectives
    ➔ Analytic and Info Requirements
    ➔ Project Success Criteria

☐ Consolidated findings document
Consolidated Requirements Findings

- Executive Overview
- Project Overview
- Business Requirements
  - Business Process Opportunities (matrix rows)
    - Description
    - Typical questions
    - Data feasibility
- Preliminary Data Warehouse Bus Matrix
- Success Criteria
Post-Interview: Facilitation for Next Steps

- Session with Business and IT management
  - Confirm findings
  - Get commitment
  - Prioritize

- For each requirement
  - Business impact / value
  - Feasibility

- Outcomes:
  - “Right” opportunities
  - Consensus
  - Ownership
  - Roadmap for growth
Linking Data Warehouse Bus Matrix and Prioritization Grid

Data Warehouse Bus Matrix

<table>
<thead>
<tr>
<th></th>
<th>Dim1</th>
<th>Dim2</th>
<th>Dim3</th>
<th>Dim4</th>
<th>Dim5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biz Proc A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Biz Proc B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Biz Proc C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Biz Proc D</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Biz Proc E</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Potential Business Impact

- Low
- High

Low → Feasibility → High

© 2003 Kimball Group  All rights reserved.
Tips for Handling Obstacles

“Abused User” ➔ Review documentation “Validate” earlier input Use alternative forum

“Overbooked User” ➔ Get sponsor’s help Defer, if prevalent

“Comatose User” ➔ Don’t prolong agony Find replacement

“Overzealous User” ➔ Assess homogeneity Reschedule sessions

“Non-Existent User” ➔ Typically fatal - STOP!
Business Requirements

Warning Signs

- “We already understand the requirements -- better than the users do.”

- “We’ll send out a survey.”

- “We’ll just look at their current reports and data files to understand users’ requirements.”

- “Since each interview lasts one hour, we’ve scheduled 2 days of interviews with 8 interviews per day.”

- “We don’t have time to document the interviews.”
Business Requirements

Summary

- Understanding business requirements is CRITICAL to successful data warehousing
- Don’t overlook the up-front preparation
- Focus on listening
- Document what you’ve heard
- Close the loop following requirements