

TECHNOLOGY TRANSFER PRESENTS

# MICHAEL **SCHMITZ**

---

**Optimizing ENTERPRISE  
DATA WAREHOUSE Design  
Utilizing DIMENSIONAL  
NORMAL FORM**

**JUNE 7-9, 2010**

---

**ETL for the  
Data Warehouse:  
a Template-Driven  
Approach**

**JUNE 10-11, 2010**

VISCONTI PALACE HOTEL - VIA FEDERICO CESI, 37  
ROME (ITALY)



info@technologytransfer.it  
www.technologytransfer.it

## ABOUT THIS SEMINAR

Dimensional Normal Form is a new approach to Data Warehouse Data Architecture which combines the strengths of both the normalized and the dimensional design paradigms to provide usable, flexible, scalable, and high performing schemas for the Enterprise Data Warehouse. Dimensional Normal Form allows an Enterprise Data Warehouse to be constructed a Data Mart at a time without requiring an intermediate Data Warehouse. While this sounds like the Kimball approach it is differentiated by its completely normalized ETL dimensions and its methods for tracking dimension history and differentiating contextual from detail or audit history.

This seminar fully covers design techniques for Data Warehousing and BI solutions based on the Dimensional Normal Form data architecture and discusses the pros and cons of the many design decisions that must be made. History considerations are discussed in detail along with their impact on schema design.

Various schema design examples are presented and discussed. The seminar participants will also be given design exercises will have their solutions analyzed and discussed by the group.

The seminar will also present contrasting Data Warehouse architectures and discuss the variations in physical design that are required for different Data Warehousing environments.

The following seminar on Template-Driven ETL will show how to build, populate and maintain the schemas built following the Dimensional Normal Form Data Architecture.

### 1. Preliminary Discussion

- Data Warehousing/Business Intelligence
- Business Drivers
- Implementation Topologies
- Keys to Success

### 2. Data Modeling Review

- Guiding Principles
- Types of Models
- Normalization

### 3. The Dimensional Approach

- The Data Warehouse Design Challenge
- Dimensional Approach History
- Dimension Table Basics
- Dimensional Table Traits, Types, and Examples
- Dimension Table Families
- Fact Table Traits, Types, and Examples
- Fact Table Granularity, Dimensionality, & Cubes
- Fact Table Content, Measures and Additively

### 4. Dimensional Normal Form

- Dimensional Normal Form
  - ETL Structures
  - Query Structures
- Conversion from 3NF to DNF
- Conversion Design Exercise

### 5. Planning BI Database Design

- Identify and Prioritize the Processes
- Prioritization Matrix
- Requirements Gathering
- Source System Analysis
- Preliminary Design
- Refinement
- Dimension Families

- Dimension Attribute History
- Audit & Data Quality Structures
- Initial Summary Strategy

## 6. Business Dimensions in Depth

- The Date Dimension Family
- The Time Dimension
- Condition or Causal Dimensions
- Separate Business Dimensions with the Same Entity Base
- Why not to Snowflake the Query Schema
- Data Warehouse Dimension Keys
- Data Quality Enforcement and Reporting
- Design Exercises

## 7. Designing for Extensibility

- The Conformed Dimension Bus
- Responding to business changes

## 8. Enterprise Data Warehouse Modeling Case Study

## 9. Tracking Dimension Attribute History

- Methodology
- Considerations
- Dimension Attribute History Worksheet
- Methods, Decision Matrix
- Case Study

## 10. Additional Design Issues

- Using Embedded Dimensions for Performance and Usability
- Heterogeneous Dimensions and Fact Tables / Vertical Partitioning
- Hierarchically Related Fact Tables
  - Consolidation
  - Surrogate Fact Table Keys
- Dimension Attributes, Levels, and Hierarchies

- Degenerate Dimensions
  - Transaction Detail Fact Tables
  - Transaction Summary Fact Tables
- Handling Large Dimensions
  - Analytical (Profile) Dimensions
  - Correlation Dimensions
  - Dimensions acting as Fact Tables
- Multi-valued Dimension to Fact Table Relationships
  - Mix Dimensions
  - Bridge Tables
  - Associated Fact Tables
  - Concatenated String Dimensions
- Value Range Dimensions
- Value Band Table
- Bracketing Attributes
- Variable-Depth Hierarchies
- Supporting One to Many Dimension to Dimension Member Relationships
- Supporting Custom Groups
- Enabling Percentage Calculations
- Nulls and Dimension Tables
- Multiple Units of Measures
- Distributed Dimensions

## 11. Real-Time (?) BI

## 12. Multi-National Support

- Situations
- Issues
- Multi-National Calendars
- Time Zones
- Multiple Currencies

## 13. Summary Strategies and Implementation

- Considerations
- Selection
- Sort Order Impact
- Automatic Summary Table Navigation and Maintenance
- Summary Selection Exercise

## 14. Physical Design and Performance Tuning

- Naming Conventions
- Storing Derived Measures
- Implementing Count Columns
- Partitioned Fact Tables
- Partitioning Transaction Detail Dimension Tables
- Index Types
  - Btree Indexes
  - Bitmap Indexes
- Dimensional Query Optimization
  - Pair-Wise Joins
  - Composite Index
  - Index Intersection
- Fact Table Indexing
- Dimension Table Indexing
- Performance Tuning

## 15. DW/BI Architecture Topics

### WHO SHOULD ATTEND

- BI and DW Managers and Project Leaders
- BI Architects, Designers and Developers
- ETL Architects, Designers and Developers
- Data Warehouse DBAs and Data Base Developers
- Managers seeking to understand the intricacies of BI Support

## **ABOUT THIS SEMINAR**

Extract, transformation, and load process development (ETL) typically accounts for more than half of the work on a Data Warehouse project. Although complex and challenging a rigorous ETL process ensures Data Quality and currency thus ensuring Data Warehouse credibility and usefulness. The good news is that using a standardized approach along with proven techniques and templates can exponentially lessen the amount of effort required and can ensure data quality, scalability and performance. This class gives a broad overview of ETL processing for the Data Warehouse and delves into the in-depth issues and considerations involved. The class looks at the increasing need for Real-Time data feeds to the Warehouse and discusses the various methods to meet these needs. It specifically presents and teaches a Template-Driven approach which quickens development speed and provides completeness. These templates are demonstrated with working Informatica/Oracle code, but can and have been adapted for other ETL tools and database platforms. They are also applicable for hand-coded efforts. CD's containing the code will be distributed to all class attendees.

Topics include:

- Basic ETL Overview
- Current ETL Architectures
- ETL Tools and Hand-coding Considerations
- In-Depth ETL Processing Techniques
- Near Real-Time Methods
- Standardized ETL Maintenance Templates

### **WHAT YOU WILL LEARN**

- Gain a thorough understanding of the critical ETL development issues
- Understand current main-stream ETL architectural approaches
- Learn in-depth techniques for addressing common development issues including how to develop near Real-Time data feeds
- Be introduced to using standardized maintenance templates and learn how to apply them to your particular environment
- Take back working code to jump start your ETL development efforts

<p><b>1. Preliminary Discussion</b></p> <ul style="list-style-type: none"> <li>• What is ETL</li> <li>• Different Acronyms</li> <li>• Mindset</li> <li>• Major Design Objectives</li> </ul> <p><b>2. ETL Overview</b></p> <ul style="list-style-type: none"> <li>• Simplistic ETL Process Parts</li> <li>• Data Acquisition</li> <li>• Data Validation and Transformation</li> <li>• Reconciliation with Source System</li> <li>• Data Publishing</li> <li>• Auditing</li> <li>• ETL Metadata</li> </ul> <p><b>3. ETL Infrastructure and Architecture</b></p> <ul style="list-style-type: none"> <li>• Business Drivers</li> <li>• Technology Drivers</li> <li>• Building Your Architecture</li> <li>• Infrastructure Architecture</li> <li>• ETL Processing Approaches             <ul style="list-style-type: none"> <li>- ETL Tools</li> <li>- Custom Coding</li> <li>- 3rd Party Utilities</li> <li>- Specialized Packages</li> </ul> </li> <li>• ETL Architecture Specification</li> <li>• High Level Architectural Examples</li> <li>• Hardware Addendum</li> </ul> <p><b>4. History Management</b></p> <ul style="list-style-type: none"> <li>• Dimension Attribute History</li> <li>• What is Dimension Attribute History</li> <li>• Why is it important</li> <li>• Our Approach</li> <li>• Analytical and Detail Attributes</li> <li>• Analytical vs. Audit Requirements</li> <li>• Business Questions Requiring Contextual History</li> <li>• Audit Based Requirements</li> <li>• The Dimension History Worksheet</li> </ul>	<p><b>5. Metadata Driven ETL Development</b></p> <ul style="list-style-type: none"> <li>• Model-Driven Design of ETL Processes</li> <li>• Required Metadata</li> <li>• Sample Metadata Repository</li> <li>• The Reality</li> </ul> <p><b>6. Dimension Processing and Templates</b></p> <ul style="list-style-type: none"> <li>• Two Dimension Processing Situations</li> <li>• Single State Extract Processing</li> <li>• Multiple State Extract Processing</li> <li>• Templates for Single State Maintenance</li> <li>• Templates for Multi State Maintenance</li> <li>• Informatica Templates Overview</li> <li>• Processing Flow for Multiple Dimension Transaction Extracts</li> </ul> <p><b>7. Fact Table Processing and Templates</b></p> <ul style="list-style-type: none"> <li>• Dimension Key Lookup Methods</li> <li>• Divide and Conquer</li> <li>• Mix Methods</li> <li>• Fact Table Process Diagram</li> <li>• Ordering Data in Fact Tables</li> <li>• Summary Table Strategy and Implementation</li> </ul> <p><b>8. Supporting Real-Time Requirements</b></p> <ul style="list-style-type: none"> <li>• What is your Requirement?</li> <li>• A Business Example</li> <li>• The Real-Time Requirement</li> <li>• A Case Study</li> </ul>	<p><b>WHO SHOULD ATTEND</b></p> <ul style="list-style-type: none"> <li>• Data Warehouse Architects, Designers and Developers</li> <li>• ETL Architects, Designers and Developers</li> <li>• Data Warehouse DBAs</li> <li>• Managers seeking to understand the intricacies of ETL Development</li> <li>• Business Intelligence and OLAP designers and developers working with ETL</li> </ul>
--	---	---

# INFORMATION

<p><b>PARTICIPATION FEE</b></p> <p><b>Optimizing Enterprise Data Warehouse Design Utilizing Dimensional Normal Form</b> € 1500</p> <p><b>ETL for the Data Warehouse: a Template-Driven Approach</b> € 1200</p> <p><b>Special price for the delegates who attend both seminars</b> <b>€ 2500</b></p> <p>The fee includes all seminar documentation, luncheon and coffee breaks.</p> <p><b>VENUE</b></p> <p>Visconti Palace Hotel Via Federico Cesi, 37 Rome (Italy)</p>	<p><b>SEMINAR TIMETABLE</b></p> <p>9.30 am - 1.00 pm 2.00 pm - 5.00 pm</p> <p><b>HOW TO REGISTER</b></p> <p>You must send the registration form with the receipt of the payment to: <b>TECHNOLOGY TRANSFER S.r.l.</b> Piazza Cavour, 3 - 00193 Rome (Italy) Fax +39-06-6871102</p> <p><b>within</b> <b>May 24, 2010</b></p> <p><b>PAYMENT</b></p> <p>Wire transfer to: Technology Transfer S.r.l. Banca Intesa Sanpaolo S.p.A. Agenzia 6787 di Roma Iban Code: IT 34 Y 03069 05039 048890270110</p>	<p><b>GENERAL CONDITIONS</b></p> <p><b>GROUP DISCOUNT</b> If a company registers 5 participants to the same seminar, it will pay only for 4. Those who benefit of this discount are not entitled to other discounts for the same seminar.</p> <p><b>EARLY REGISTRATION</b> The participants who will register 30 days before the seminar are entitled to a 5% discount.</p> <p><b>CANCELLATION POLICY</b> A full refund is given for any cancellation received more than 15 days before the seminar starts. Cancellations less than 15 days prior the event are liable for 50% of the fee. Cancellations less than one week prior to the event date will be liable for the full fee.</p> <p><b>CANCELLATION LIABILITY</b> In the case of cancellation of an event for any reason, Technology Transfer's liability is limited to the return of the registration fee only.</p>
--	---	--

## MICHAEL SCHMITZ

**OPTIMIZING ENTERPRISE DATA WAREHOUSE DESIGN UTILIZING DIMENSIONAL NORMAL FORM**

Rome June 7-9, 2010  
Visconti Palace Hotel - Via Federico Cesi, 37  
Registration fee: € 1500

**ETL FOR THE DATA WAREHOUSE: A TEMPLATE-DRIVEN APPROACH**

Rome June 10-11, 2010  
Visconti Palace Hotel - Via Federico Cesi, 37  
Registration fee: € 1200

**BOTH SEMINARS**

Special price for the delegates who attend both seminars: € 2500

If registered participants are unable to attend, or in case of cancellation of the seminar, the general conditions mentioned before are applicable.

first name .....

surname .....

job title .....

organisation .....

address .....

postcode .....

city .....

country .....

telephone .....

fax .....

e-mail .....



Stamp and signature

Send your registration form with the receipt of the payment to:  
**Technology Transfer S.r.l.**  
Piazza Cavour, 3 - 00193 Rome (Italy)  
Tel. +39-06-6832227 - Fax +39-06-6871102  
info@technologytransfer.it  
www.technologytransfer.it



## SPEAKER

**Michael Schmitz** specializes in high performance Data Warehousing solutions. He has in-depth experience utilizing efficient, scalable techniques whether dealing with large-scale Data Warehouses or small-scale, platform constrained Data Mart implementations. After establishing a career in database management systems and working for Software AG and IBM he began specializing in Data Warehousing in 1987 with Teradata. He subsequently worked for Red Brick and Informix and served as Director of Data Warehousing for Oracle's Mid-Market Solutions Group. After establishing a successful independent consultancy he worked Knightbridge Solutions and HP's Information Management practice. He is most recently acted as the Lead DW/BI Architect for Nike building a new Enterprise Data Warehouse based on Informatica and Teradata. He is currently the principal and owner of Database Performance LLC. His experience with different systems, architectures, and methodologies has enabled him to evolve a unique approach to Data Warehousing utilizing best of breed components and methods. He has written several articles on his approach and has co-authored one book. He actively works on Data Warehouse implementations and teaches design and implementation classes internationally.