tdwi
Transforming Dat
With Intelligence

Offered by **Chris Adamson** through **TDWI Onsite Education**
The companion course to the book *Star Schema: The Complete Reference.*

Course Agenda, Day 1 of 2

Module One
Fundamentals

- Dimensional Modeling
   o Process measurement
   o Uses beyond database design
- Information Architectures and Dimensional Data
   o Inmon's CIF Architecture
   o Kimball's Dimensional Bus
   o Stand-alone Data Marts
- Fundamental Terms and Best Practices
   o Facts and dimensions
   o Surrogate keys, natural keys and slow change processing
   o Fundamental concepts including grain, sparsity and additivity
   o Best practices for fact tables and dimension tables
- Implementations
   o Relational (star and snowflake)
   o Multidimensional (cube)
   o Cubes as primary store
   o Cubes as derived data stores

Module Two
Multiple Stars

- Multiple star solutions
- Designing multiple fact tables
   o Identifying multiple processes
   o Differences in dimensionality
   o The pitfalls of single fact table design
- Using multiple stars
   o How not to query multiple fact tables
   o The concept of drilling across
   o What you need to know about your query and reporting tools
- Conformance and business value
   o High impact business questions span processes
   o The concept of conformance
   o Ensuring subject areas work together
   o Enabling incremental implementation

Module Three
Advanced Fact Table Design

- Transaction schemas
  - Transaction grain
  - Shortcomings of transaction designs
- Periodic Snapshots
  - Snapshot grain and period
  - Semi-additivity, density, and impact on BI
  - Building both transaction and snapshot schemas
  - Snapshots and averages
- Accumulating Snapshots
  - Studying process efficiency
  - Accumulating metrics in a single row
  - Lag analysis
  - Impacts on slow change processing and data integration
  - Building both transaction and accumulating snapshots
- Factless Fact Tables
  - Processes that seem to lack metrics
  - Factless fact tables that track events
  - Pros and cons of adding constant-value fact
  - Factless fact tables that track conditions
  - Comparing conditions to actual events
- Heterogeneous Attributes
  - Attributes that vary based on category
  - The impact of modeling a single set of attributes
  - Core and custom dimensions
  - Core and custom fact tables
  - Alternatives to core and custom solutions

Module Four
Design and Data Integration

- The data integration process
  - Loading a dimension table
  - Loading fact tables
  - Key lookup processing
- Designing to aid key management
  - The impact of slow changes on key lookups
  - Adding attributes to aid lookup process
- Designing to reduce slow change bottlenecks
  - The bottlenecks of slow change processing
  - How dimension checksums aid slow change processing
- Specifying data transformation rules
  - How to capture standard translations and data quality rules
  - Why this must be done at design-time
- Invalid or late reference data
  - Receiving valid facts with invalid or missing reference data
  - Adding rows to dimension tables so that facts can be loaded
  - Adjusting facts when reference data arrives

- Adding columns to support QA
  - Housekeeping columns in dimension tables
  - Housekeeping dimensions for fact tables

Course Agenda, Day 2 of 2

Module Five
Advanced Dimension Design

- Understanding hierarchies
  - Drilling with and without hierarchies
  - Multiple hierarchies in one dimension
  - Impact of hierarchies on BI, ETL and DBA perspectives
  - Why to model hierarchies
  - Snowflake schemas
- Dimension Reuse
  - Modeling roles
  - Querying with roles
- Nulls
  - Problems introduced by NULL values
  - Avoiding NULL dimension attributes
  - Avoiding NULLs with optional relationships
  - NULL facts

Module Six
Advanced Slow Change Processing

- Time-stamped Dimensions
  - Why type 2 is not good enough
  - Time-stamped dimensions (transaction dimensions)
  - Using with a fact table
  - Hybrid attributes that behave like facts and dimensions
- Mini-dimensions (Types 4 & 5)
  - Large and expanding dimension tables
  - The use of a mini-dimension to stem growth
  - Impact on schema capability
  - Type 4 and Type 5 slow changes
  - Loading the mini-dimension
- Current and Previous (Type 3)
  - Limited access to unchanged value
- Hybrid Responses
  - Tracking both Type 1 and Type 2 responses
  - For dimension Attributes (Type 6)
  - For entire dimension tables (Type 7)

Module Seven
Bridge Table Design

- The dimension bridge
  - When a single fact needs to reference more than one dimension row

---

- o The "flattening" option
- o The bridge table
- o Use of the bridge
- o Avoiding double counting with an allocation factor
- o Avoiding double counting by hiding the bridge (sandbox)
- • The attribute bridge
  - o A dimension attribute repeats for a single dimension row
  - o The attribute bridge
  - o Use of the bridge and avoidance of double-counting
  - o Bridge table vs. factless fact table
- • The hierarchy bridge
  - o Recursive relationships (ragged hierarchies, unbalanced hierarchies, variable depth hierarchies and instance hierarchies)
  - o Flattening and backfilling
  - o The hierarchy bridge and its structure
  - o Using the bridge to "roll up" a recursive hierarchy
  - o Using the bridge to "roll down" the hierarchy
  - o ETL implications of hierarchy change and slow changes

Module Eight
Scaling Dimensional Designs

- • Scaling and scope
  - o Conformance across subject areas
  - o Conformed dimensions that are not identical
  - o Conformed rollups and overlapping dimensions
  - o An advanced conformance matrix
- • Derived schemas and performance
  - o Merged fact tables
  - o Pivoted fact tables
  - o Set operations on fact tables
  - o Sliced or partitioned fact tables
- • Aggregate schemas and performance
  - o N-way aggregate design
  - o Aggregate portfolio and impact on throughput
  - o Aggregates and aggregate navigation

Module Nine
Design Tasks and Deliverables

- • Design and data architecture
  - o Dimensional design as architecture task
  - o Architecture and incremental implementation
  - o Implementation without up-front architecture activity
- • Design activities
  - o Key design activities
  - o Resources required for design success

- o Estimating duration of design activities
- o Requirements, top level design, detailed design and roadmap
- Documenting requirements (Conceptual design)
  - o Subject area definition
  - o Metric groups and metric definition
  - o Business hierarchies
  - o Conformance
- Top level design
  - o Key elements of dimension table design
  - o Key elements of fact table design
  - o Documenting what is not visible
  - o The conformance matrix
  - o The design review
- Detailed design
  - o Attribute level design of fact and dimension tables
  - o Technical design elements (DBA perspective)
  - o Business definitions (business perspective)
  - o Source to target mappings (integration perspective)
- Roadmap
  - o Project definitions
  - o Managing scope of implementations
- Implementation approaches
  - o Time-boxed projects with iteration
  - o Agile projects (with and without up-front discovery)
  - o Typical team member roles for implementation projects

Appendix A
Exercises

- Ten modeling exercises designed to reinforce concepts of each module
  - o Two are completed during the class (one each day)
  - o Students can work through remaining exercises on their own
- Exercise sessions
  - o Students are divided into groups of 4-5 to complete the exercises
  - o Flip chart or white board must be provided for each group
  - o Exercise is discussed by class as a whole upon completion

Appendix B
Exercise Solutions

- o Solution pages for each exercise
- o Example models and discussion points