Previews of TDWI course books offer an opportunity to see the quality of our material and help you to select the courses that best fit your needs. The previews cannot be printed.

TDWI strives to provide course books that are content-rich and that serve as useful reference documents after a class has ended.

This preview shows selected pages that are representative of the entire course book; pages are not consecutive. The page numbers shown at the bottom of each page indicate their actual position in the course book. All table-of-contents pages are included to illustrate all of the topics covered by the course.
<table>
<thead>
<tr>
<th>Module</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 1</td>
<td>Data Modeling Concepts</td>
<td>1-1</td>
</tr>
<tr>
<td>Module 2</td>
<td>Business Data Models</td>
<td>2-1</td>
</tr>
<tr>
<td>Module 3</td>
<td>Logical Data Models</td>
<td>3-1</td>
</tr>
<tr>
<td>Module 4</td>
<td>Implementation Data Models</td>
<td>4-1</td>
</tr>
<tr>
<td>Module 5</td>
<td>Summary and Conclusion</td>
<td>5-1</td>
</tr>
<tr>
<td>Appendix A</td>
<td>Entity-Relationship Modeling Basics</td>
<td>A-1</td>
</tr>
<tr>
<td>Appendix B</td>
<td>The Data University Case Study</td>
<td>B-1</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Exercises</td>
<td>C-1</td>
</tr>
<tr>
<td>Appendix D</td>
<td>Bibliography and References</td>
<td>D-1</td>
</tr>
</tbody>
</table>
To learn:

✓ Differences in modeling techniques for business transactions, business events, and business metrics
✓ Different types of data and their implications
✓ Application of business context to modeling activities
✓ The role of business requirements in BI data modeling
✓ The role of source data analysis in data modeling
✓ Use of normalized modeling techniques for data warehouse analysis and design
✓ Use of dimensional modeling techniques for data warehouse analysis and design
✓ The roles of generalization and abstraction in data warehouse design
✓ The roles of identity and hierarchy management in master data management
✓ How time-variant data is represented in data models
✓ Implementation and optimization considerations for warehousing data stores
Module 1

Data Modeling Concepts

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Data Modeling Life Cycle</td>
<td>1-2</td>
</tr>
<tr>
<td>Kinds of Data Systems</td>
<td>1-6</td>
</tr>
<tr>
<td>Data Taxonomy</td>
<td>1-8</td>
</tr>
<tr>
<td>Data Modeling Framework for BI</td>
<td>1-12</td>
</tr>
</tbody>
</table>
The Data Modeling Life Cycle
Where Data Modeling Begins and Ends

Business Activities & Information Needs

Business Operations
Business Planning
Business Performance Management

Business Transaction Systems
Decision Systems
Business Analytics Systems

Business Data
Application Data Files & Tables
Data Warehouse
Operational Data Store (ODS)
Relational Data Marts
Reporting Databases
Reporting Flat Files
Published Reports

OLAP Data Marts
Data Mining Files
Analytics Applications
The Data Modeling Life Cycle
Where Data Modeling Begins and Ends

THE BEGINNING
A data model is an abstract representation of the data in an enterprise, or of the information that is derived from that data. Data and information can (and should) be represented at multiple levels of abstraction, each providing a different perspective and understanding of the data. The highest level of abstraction is a business context view with both external (outside looking in) and internal (looking from within) perspectives. Thus data modeling begins with business activities and the information needs of those activities. This view describes the scope of and the context for business information requirements—a sensible start to modeling the right data.

THE END
The natural conclusion of data modeling is implemented data—data files and database tables. Far from the top level of abstraction, implemented data is beyond the bottom tier; it is no longer abstract but real and physical. At this level attention turns to the right implementation for the data.

THE MIDDLE
The complexities and challenges of data modeling lie between the top layer of context and the bottom tier of implementation. Getting from the right data to the right implementation involves an understanding of the business and many information systems, ranging from those that capture data in the course of business activities to those that turn data into information and supply that information to the business.
Kinds of Data Systems

Business Uses of Data

- Business Operations
- Business Planning
- Business Performance Management

Business Activities & Information Needs

- Business Transaction Systems
- Decision Systems
- Business Analytics Systems

Business Context (External View)

Business Concept (Internal View)

Business Design (Business Systems View)

Technical Design (Computer Systems View)

Technical Specification (Technology View)

Business Data

- Application Data Files & Tables
- Data Warehouse
- Operational Data Store (ODS)
- Relational Data Marts
- Reporting Databases
- Reporting Flat Files
- Published Reports
- OLAP Data Marts
- Data Mining Files
- Analytics Applications
Kinds of Data Systems
Business Uses of Data

MODELING BASED ON BUSINESS USE

Multiple business uses, and multiple kinds of information systems to support those uses, add complexity in the progression from information requirements to data implementations. Although they work with the same scope of data and information, the designs and databases must adapt to serve multiple masters. They must meet the needs of:

- Business Transaction Systems – Day-to-day operational systems used to record and report transactions that occur in the course of doing business. These include customer transactions as well as financial, inventory, supplier, employee, and other kinds of business transactions.

- Decision Systems – Query, reporting, and data access systems that provide the information services that support the tactical analysis and decision processes of the business. Ad hoc query, managed query, demand reporting, and published reports are all examples. Note that this is not the standalone decision support systems (DSS) of old which combine reporting with analytic systems and are typically non-integrated solutions for a discrete community of users.

- Business Analytics Systems – Business measurement systems that support the need to monitor and manage performance both strategically and tactically. Performance dashboards, business scorecards, and online analytic processing are examples. These systems provide the business metrics and measures capabilities of older DSS systems with expanded capability, and ideally with greater integration across the enterprise.
# Module 2

Business Data Models

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Context</td>
<td>2-2</td>
</tr>
<tr>
<td>Business Data Model Development</td>
<td>2-14</td>
</tr>
<tr>
<td>Gathering Business Questions</td>
<td>2-16</td>
</tr>
<tr>
<td>Analyzing Business Questions</td>
<td>2-24</td>
</tr>
<tr>
<td>Fact Analysis and Refinement</td>
<td>2-36</td>
</tr>
<tr>
<td>Qualifier Analysis and Refinement</td>
<td>2-40</td>
</tr>
<tr>
<td>Fact/Qualifier Analysis Results</td>
<td>2-42</td>
</tr>
<tr>
<td>Business Dimensional Modeling</td>
<td>2-44</td>
</tr>
</tbody>
</table>
Business Context
Business Drivers, Goals, and Strategies

Business Drivers
forces that cause a need to act

Business Goals
desired outcomes of actions

Business Strategies
planned means to achieve goals

forces that cause a need to act:
• increased competition
• changing regulations
• changing economy
• merger or acquisition
• changing marketplace
• changing workforce
• changing technology
• etc.

desired outcomes of actions:
• increased market share
• increased customer share
• improved customer retention
• cost reduction
• increased revenue
• improved profit margins
• reduced cycle time
• etc.

planned means to achieve goals:
• competitive pricing
• innovative product packaging
• customer loyalty programs
• outsourcing and partnerships
• new sales channels
• supply chain optimization
• new delivery channels
• etc.
Business Context
Business Drivers, Goals, and Strategies

THE MODELING FRAMEWORK

WHY MODEL
Business context determines the nature of data and information services – the business processes to be affected, the kinds of applications to be implemented, and the information services to provide. Business context provides the means to align data with business goals. The contextual representation is generally determined prior to initiating a project. Project team members should be aware of and use the context to establish a more complete and relevant set of requirements.

BUSINESS DRIVERS
Business drivers are those things that are strategically important in positioning the business to achieve its short- and long-term goals. They are the external forces that have significant influence on operation and performance of a business. Drivers create need to take action, but they don’t dictate the actions to be taken. Common business drivers include changing economic, political, social, and technological factors.

BUSINESS GOALS
Business goals are the things that the business wants to accomplish to respond to business drivers. Drivers create the need to act. Goals describe the desired outcomes of taking action. Goals are commonly related to financial or operational performance (i.e., cost reduction, generation of revenue, increased market share, etc.). Goals are most effective in setting data management priorities and directions when they are: (1) described by clear, concise, understandable statements, (2) specific enough that level of achievement can be measured, and (3) of high business priority.

BUSINESS STRATEGIES
Business strategies are plans to turn the defined business goals into reality. They include the things that the business will do to shape its future and achieve the business goals. The range of strategies is broad – introducing new products, exploiting new sales channels, pricing competitively, optimizing business processes, etc. Strategies help to determine which business processes and organizations most need to be information enabled.
Business Data Model Development
Top-Down – Incremental and Iterative
Business Data Model Development
Top-Down – Incremental and Iterative

THE MODELING FRAMEWORK

WHY MODEL
The business data model is a view of the major things (i.e., entities) of interest from a business perspective, independent of other factors such as processes, organizational structure, or technology. While it is not physically implemented, applied properly, it enables enterprises to create consistent data structures. It is also sometimes known as the “enterprise data model” or “conceptual data model”.

PROCESS
The recommended process consists of the following major activities:

- Determine the scope.
- Determine the entities of interest, and establish a business-oriented name and definition for each.
- Determine relationships among entities, and establish the verb phrase describing the relationship, the optionality, and the cardinality for each.
- Determine the attributes of interest, and establish a business-oriented name and definition for each.
- Determine appropriate generalizations and specializations.
- Normalize the model to at least third normal form.
Module 3
Logical Data Models

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>What to Model</td>
<td>3-2</td>
</tr>
<tr>
<td>Understanding Data Sources</td>
<td>3-4</td>
</tr>
<tr>
<td>Logical Relational Modeling</td>
<td>3-10</td>
</tr>
<tr>
<td>Logical Dimensional Modeling</td>
<td>3-20</td>
</tr>
<tr>
<td>Logical Models and Business Metrics</td>
<td>3-34</td>
</tr>
<tr>
<td>Logical Models and Business Analytics</td>
<td>3-40</td>
</tr>
<tr>
<td>Logical Models and Master Data Management</td>
<td>3-44</td>
</tr>
<tr>
<td>Logical Models and Unstructured Data</td>
<td>3-48</td>
</tr>
</tbody>
</table>
What to Model

The Data and Information Pipeline

source data

staged data

warehouse data

user views

Multi-Dimensional Data Cubes

Database Tables

DataMarts

Business Information

End-User Data

Flat Files

Extract

Extract

Extract

Transform

Transform

Transform

Load File

Load File

Load Files

Load Files

Generate Cube

Load Database

Load Database
What to Model
The Data and Information Pipeline

In a typical data warehousing environment data flows from sources, through a sequence of intermediate processes and data stores, to be delivered ultimately as information that is useful to business. The data modeler’s challenge – with data stores including sources, staging, data warehouses, and data marts – is to determine which data to model and at what level to model the data.

Every project must determine the right amount of data modeling to be done. Compromises are often necessary to achieve appropriate balance between data complexity and the real project constraints of time and resources. A practical minimum standard includes:

- User view modeling that is sufficient to understand end-user requirements for data.

- Warehouse modeling that is detailed enough to achieve data integration and to serve as ongoing documentation of the data warehouse.

- Source data modeling as needed to understand the content and structure of data sources.
Logical Models and Business Metrics
Creating a Catalog of Metrics

[Diagram showing categories such as Organization, Enterprise, Activity, Process, Management Disciplines, CRM, BPM, SCM, BAM, etc., with Business Domains including Education, Research, Fund-Raising, Info Mgt, etc.]
Logical Models and Business Metrics
Creating a Catalog of Metrics

**THE MODELING FRAMEWORK**

**WHY MODEL**

Measurement-based disciplines have become central to management of business. BPM, SCM, CRM and more take business-by-the-numbers to a new level. Technology enables the trend with performance scorecards and dashboards. But with these technologies comes the risk of new integration problems – disintegration of business analytics. How do we prevent the customer measures of CRM from conflicting with those of SCM? How do we achieve consistency, cohesion, and integration among metrics?

If we’ve learned anything from the past we will start with a business view and enterprise perspective. But what really happens when we model business measures? Most of us start with a star-schema goal – not because we don’t want a better way, but because it is the only thing that is widely known and accepted. Go to the internet and Google the phrase “business metrics model.” As of this writing, very few hits contain all three words; and those are simply coincidental uses of the terms, not information about modeling of business metrics.

**WHAT TO MODEL**

This modeling activity produces a catalog of implemented and desired business metrics with consistency of description and classification for all metrics. The catalog supports multiple dimensions to organize metrics by

- community of interest (enterprise, organization, process, etc.)
- business domain (mission or resource domain)
- external drivers (economic, social, political, technological)
- internal perspectives (finance, customer, process, people)
- management discipline (BPM, CRM, SCM, etc.)
## Module 4
Implementation Data Models

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Structure in Transaction Systems</td>
<td>4-2</td>
</tr>
<tr>
<td>Structural Modeling and Data Integration</td>
<td>4-4</td>
</tr>
<tr>
<td>Structural Modeling and Business Analytics</td>
<td>4-26</td>
</tr>
<tr>
<td>Physical Design Overview</td>
<td>4-44</td>
</tr>
<tr>
<td>Some Optimization Techniques</td>
<td>4-46</td>
</tr>
<tr>
<td>Physical Design and Implementation</td>
<td>4-58</td>
</tr>
</tbody>
</table>
Structural Modeling and Data Integration
From Business Models to Technology Models
Structural Modeling and Data Integration
From Business Models to Technology Models

THE MODELING FRAMEWORK

WHY MODEL
The logical data model for the data warehouse or ODS is a business-oriented model. It is a business system design that is free of data system dependencies. Those dependencies, however, must be recognized and resolved before physical design and implementation occur. That is the role of structural modeling.

WHAT TO MODEL
Information systems dependencies occur in warehouse and ODS data for two primary reasons:

- Time-variance is a defining characteristic of both the data warehouse and the ODS.
- Access, navigation, and security requirements are significant influences on warehouse and ODS implementation.

Both reasons are cause to compromise the rules of normalization when moving from logical design toward physical design and implementation. Although best-practice logical models are normalized to a defined level, normalization is also a structural design issue for several reasons. It is important to ensure that the logical model is in the desired form. It is equally important to understand where and why that model deviates from the third normal form. This understanding is essential foundation before further compromises of normalization rules can be made with confidence.

WHEN TO MODEL
Develop structural models following logical modeling and prior to physical design. Many modelers consider structural modeling to be the first step of physical design.

HOW TO MODEL
Who: This is a job for data modelers and the development team.
Inputs: The logical data model is the primary input. Knowledge of how the data warehouse or ODS will be used is also important.
Outputs: This process produces an entity-relationship model that is adapted to the requirements of time-variance, access, navigation, and security. It is frequently considered to be the first-cut physical model.
Activities: The activities of adjusting for time, access, navigation, and security are discussed in greater detail on the following pages.
Physical Design Overview
The Results of Physical Design and Implementation
Physical Design Overview
The Results of Physical Design and Implementation

**DESIGN MEETS TECHNOLOGY**

The diagram on the facing page highlights the levels of activity that constitute physical design and implementation. As previously described the logical design develops system models from a technology-neutral point of view, and structural design extends to technology-oriented models without becoming platform specific.

Physical design optimizes the structural models to work with chosen technologies – database management systems, query tools, etc. Physical design considerations include performance, size and growth, availability, recovery from failure, and use of specific technology features.

Implementation then creates the computer objects – files, tables, etc. – that constitute a data system.