TDWI Data Warehouse Automation
Better, Faster, Cheaper ... You Can Have It All
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.
COURSE OBJECTIVES

You Will Learn:

- Concepts, principles, and practices of data warehouse automation
- The current state of data warehouse automation technology
- Automation opportunities and benefits when building or managing a data warehouse
- How to get started with data warehouse automation
- Best practices and mistakes to avoid with data warehouse automation

TDWI takes pride in the educational soundness and technical accuracy of all of our courses. Please send us your comments—we’d like to hear from you. Address your feedback to:

info@tdwi.org

Publication Date: October 2017

© Copyright 2017 by TDWI. All rights reserved. No part of this document may be reproduced in any form, or by any means, without written permission from TDWI. Product and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies.
| Module 1 | Data Warehouse Automation Concepts and Principles... | 1-1 |
| Module 2 | Building and Managing the Data Warehouse............... | 2-1 |
| Module 3 | Using Data Warehouse Automation.......................... | 3-1 |
| Module 4 | Data Warehouse Automation in Action....................... | 4-1 |
| Module 5 | Getting Started with Data Warehouse Automation.......... | 5-1 |
| Appx A   | Bibliography and References.................................. | A-1 |
# Module 1

Data Warehouse Automation Concepts and Principles

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Warehouse Automation Basics</td>
<td>1-2</td>
</tr>
<tr>
<td>Why Data Warehouse Automation?</td>
<td>1-10</td>
</tr>
<tr>
<td>The Foundation</td>
<td>1-16</td>
</tr>
<tr>
<td>Activities and Deliverables</td>
<td>1-22</td>
</tr>
<tr>
<td>The Technology Landscape</td>
<td>1-26</td>
</tr>
</tbody>
</table>
Why Data Warehouse Automation?

Business Benefits

- **BETTER SOLUTIONS**
  business is able to make changes far later in the development process

- **BUSINESS AGILITY**
  ability to respond to change and new requirements quickly ... both to build new solutions and to adapt existing solutions

- **SPEED**
  faster time to build and deploy new capabilities, to adapt and modify existing warehouses, and to build and review prototypes

- **COST SAVINGS**
  reduced cost through time savings, labor savings, and better quality
Why Data Warehouse Automation?

Business Benefits

**QUALITY AND EFFECTIVENESS**
Data warehouse automation delivers quality and effectiveness through the ability to build better solutions. Better solutions are those that best meet real business requirements, and it is especially difficult to get complete and correct requirements when limited to an early phase of a linear development process. With data warehouse automation the business can make changes much later in the development process and change can occur more frequently with less disruption, waste, and rework.

**AGILE BUSINESS**
Ability to change fast and frequently extends beyond the warehouse development process. Changes that occur in business requirements can be met with quick response. Responding to change in real time and without the delay of lengthy projects is the essence of business agility.

**SPEED**
Speed is the critical factor that enables agility both for agile business and for agile development. The ability to generate quickly and to regenerate equally fast when change occurs is a fundamental automation capability.

**COST SAVINGS**
Ultimately building better, building faster, and changing quickly when needed bring substantial cost savings to data warehouse development, operation, maintenance, and evolution.
The Foundation
Components of Data Warehousing
The Foundation

Components of Data Warehousing

DATA FLOW – FROM DISPARATE DATA TO INTEGRATED INFORMATION

The facing page illustrates the core elements of data warehousing, beginning with disparate data sources at the bottom of the diagram and leading to integrated information resources at the top. The information resources are not necessarily the end of the line – they are simply the end of the data integration processes. Business value is created when they are used for reporting, business intelligence, decision making, analytics, etc.

The center of the diagram shows the processing steps to get from data sources to integrated information using two methods – data consolidation through extract, transform, and load (ETL) processing and data virtualization through a series of abstract data views. Consolidation and virtualization are often combined to optimize the flow of data. Data warehouse automation can implement both approaches independently or in a mix-and-match form.

ALL OF THE PARTS

Data warehousing encompasses many different techniques and produces many components to enable the data-to-information flow. Among these are architectural standards, data models, mappings, data transformations, database load procedures, tests and controls, events and errors, and metadata. Data warehouse automation includes capabilities to create, connect, manage, and apply all of these components.
Module 2
Building and Managing the Data Warehouse

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Data Warehousing Lifecycle</td>
<td>2-2</td>
</tr>
<tr>
<td>Building the Data Warehouse</td>
<td>2-22</td>
</tr>
<tr>
<td>Managing the Data Warehouse</td>
<td>2-30</td>
</tr>
</tbody>
</table>
Building the Data Warehouse
Requirements-Driven Development

- requirements analysis
- warehouse data modeling
- source data analysis
- source/target mapping
- acquisition design & specification
- transformation design & specification
- database load design & specification
- database build
- ETL build
- start-up data preparation & loading
- system & acceptance testing
- deployment

missed or incorrect requirements
Building the Data Warehouse

Requirements-Driven Development

A TRADITIONAL APPROACH

Requirements-driven development is the traditional process that begins with requirements gathering and proceeds in a linear fashion through analysis, modeling and design, specification and coding, and deployment. This approach is a legacy of conventional software development processes that does not work well for data warehousing. It is characterized by:

- Intensive and detailed planning
- Long development timelines
- Limited business participation early in the process
- Substantial staffing and skills demand
- Difficulty in requirements gathering and analysis
- Much cycling back to previous steps to correct errors and oversights
- Much waste and rework
- Post-deployment discovery of incorrect and missed requirements
Managing the Data Warehouse

Operations

- Extension
- Expansion
- Migration
- Modernization
- Re-architecting
- Re-sourcing
- Refactoring

- Sequencing
- Dependencies
- Scheduling
- Execution
- Verification
- Validation
- Error handling

- Tracing
- Troubleshooting
- Bug fixes
- Data correction
- Error recovery
- Upgrades
- Enhancements

- Usage
- Performance
- Availability
- Access
- Security
- Growth
- Capacity

scheduling
documentation & metadata
managed environments
validation testing
Managing the Data Warehouse

Operation

**AUTOMATION FOR OPERATIONS**

Operations comprises a core set of activities in managing a data warehouse, with attention to:

- Sequencing
- Dependencies
- Scheduling
- Execution
- Verification
- Validation
- Error Handling

Automation aids data warehouse operations with features and functions for:

- Scheduling
- Documentation and Metadata
- Managed Environments
- Validation Testing
Module 3

Using Data Warehouse Automation

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automation Use Cases</td>
<td>3-2</td>
</tr>
<tr>
<td>Case Studies</td>
<td>3-20</td>
</tr>
</tbody>
</table>
Automation Use Cases

Building a New Data Warehouse

BEFORE

Disparate Data Sources

Flat Files
Databases
Cloud Databases
Hadoop and NoSQL
ERP
finance
purchasing
HR
supply chain
CRM
Spreadsheets
Third-Party Data Sources

one-time reports
long delays
surrogate & proxy data
downloads & handcrafting
unfulfilled needs

Integrated Information Needs

AFTER

Disparate Data Sources

Flat Files
Databases
Cloud Databases
Hadoop and NoSQL
ERP
finance
purchasing
HR
supply chain
CRM
Spreadsheets
Third-Party Data Sources

warehousing processes

Data Warehouse

warehouse access

Integrated Information Needs
# Automation Use Cases

## Building a New Data Warehouse

**SCENARIO**

You need to build a new data warehouse either where none exists or as a complete replacement of an existing and dysfunctional warehouse.

**CHALLENGES**

Without automation, all of the normal data warehousing challenges exist. Source data is messy, warehouses are hard to build, they take too long to build, and they are obsolete before they are deployed.

**OPPORTUNITIES**

Automation opportunities are abundant in this scenario. You can automate everything from planning to deployment and operation using any of model-driven, discovery-driven, or data-driven approaches.
Case Studies

ABInBev

Disparate data sources and geographic locations.

Automate the information lifecycle to enable 16 years of sales history across 48 dimensions from 93 transaction data sets and draw data from 11 source systems.

Newly implemented platform enables employees to track actual performance against plan.

Regulatory reporting and audit demand quality data.

Manage in one place the structure of planning data and actual performance metrics as business structures and needs change.

Enables drill-down reporting tools to see the lowest granularity of data held in the system.

Requirement to manage both global and local brands.

Previously, 70% of analyst time was spent gathering data. Now, its 80% in favor of analysis.

Differences between and differing change schedules between local government reporting requirements.


Case Studies

ABInBev

BEVERAGE INDUSTRY COMPLEXITIES

Founded in London, Ontario, in 1847 and the proud brewer of more than 60 quality beer brands, Labatt is the leading brewery in Canada and a part of Anheuser-Busch InBev, the world’s largest brewer by volume. Since the late 1990s, Labatt has evolved from a regionally federated business into a centralized one, reporting into a global parent company, Interbrew. In 2004, Interbrew merged with AmBev, creating InBev. This merger introduced another layer into what was already a complex organizational picture. Today, Labatt also markets, distributes, and sells InBev’s global brands. This requires Labatt to manage the performance of both global and local brands and report the results to its various stakeholders.

Labatt’s data environment is very rich, consisting of typical internal data sources and a large number of external data sources. This adds complexity, particularly in sales and marketing, when comparing regional performance or combining information in consistent ways.

Labatt’s vision was to integrate planning and performance management in one environment. The Enterprise Business Intelligence initiative was designed to meet that vision, providing simple, easy access to multidimensional information for business users to manage their business and analysts to gain greater insights through more powerful analytic tools. To support this performance management challenge, Labatt turned to a data warehouse automation solution that can quickly adapt to business change.

Source: http://kalido.com/portfolio-item/abinbev/
Module 4
Data Warehouse Automation in Action

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attunity Compose in Action</td>
<td>4-2</td>
</tr>
<tr>
<td>Magnitude Software Kalido in Action</td>
<td>4-4</td>
</tr>
<tr>
<td>TimeXtender in Action</td>
<td>4-6</td>
</tr>
<tr>
<td>WhereScape in Action</td>
<td>4-8</td>
</tr>
</tbody>
</table>
Module 5
Getting Started with Data Warehouse Automation

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step by Step</td>
<td>5-2</td>
</tr>
<tr>
<td>Human Factors</td>
<td>5-4</td>
</tr>
<tr>
<td>New Horizons</td>
<td>5-6</td>
</tr>
<tr>
<td>Next Steps</td>
<td>5-8</td>
</tr>
</tbody>
</table>
Step by Step
A Process for Automation

1. Build the Business Case
2. Secure Sponsorship & Buy-In
3. Short List the Vendors & Products
4. Select & Execute Proof-of-Concept Projects
5. Select and Install Technology
6. Implement & Manage Organizational Change
7. Train & Implement Data Warehouse Automation

Deliver Fast, Cost-Effective, High-Quality Data Warehousing
## Step by Step
### A Process for Automation

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BUSINESS CASE</strong></td>
<td>Build the business case for data warehouse automation based on business benefits, not technical benefits. Focus on value-creating benefits such as speed, agility, solution quality, and cost savings.</td>
</tr>
<tr>
<td><strong>SPONSORSHIP AND BUY-IN</strong></td>
<td>Seek a sponsor who can secure the funding, resources, and political will to drive data warehouse automation. Then identify other key stakeholders and work to secure their buy-in.</td>
</tr>
<tr>
<td><strong>SHORT LIST</strong></td>
<td>Identify the candidate list of vendors and products that fit your needs, constraints, and culture. When developing the short list, also identify the criteria that you will use to make a final selection.</td>
</tr>
<tr>
<td><strong>PROOF OF CONCEPT</strong></td>
<td>Identify one or two proof-of-concept projects and ask each vendor on the short list to illustrate how they will do the work. Before executing the projects establish a baseline for comparison either by: (1) automating something that you’ve already done and have known time and cost to build it manually, or (2) by building something new for which you’ve estimated time and cost to build manually.</td>
</tr>
<tr>
<td><strong>TECHNOLOGY SELECTION</strong></td>
<td>Based on proof of concept results and previously defined selection criteria, make your technology decision and install the automation tools.</td>
</tr>
<tr>
<td><strong>ORGANIZATIONAL CHANGE</strong></td>
<td>Recognize the need for and undertake the necessary organizational changes. The processes, roles, responsibilities, and team configurations for manual data warehousing are not what you will want for automation.</td>
</tr>
<tr>
<td><strong>TRAINING AND IMPLEMENTATION</strong></td>
<td>Train people to use the automation tools, and train them to understand their new roles and responsibilities with data warehouse automation.</td>
</tr>
</tbody>
</table>