



**Transforming Data
With Intelligence™**

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.

This page intentionally left blank.



TDWI Data Warehouse Automation

Better, Faster, Cheaper ... You Can Have It All

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.

TABLE OF CONTENTS

Module 1	<i>Data Warehouse Automation Concepts and Principles...</i>	1-1
Module 2	<i>Building and Managing the Data Warehouse</i>	2-1
Module 3	<i>Using Data Warehouse Automation</i>	3-1
Module 4	<i>Data Warehouse Automation in Action</i>	4-1
Module 5	<i>Getting Started with Data Warehouse Automation</i>	5-1
Appx A	<i>Bibliography and References.....</i>	A-1



Module 1

Data Warehouse Automation Concepts and Principles

Topic	Page
Data Warehouse Automation Basics	1-2
Why Data Warehouse Automation?	1-10
The Foundation	1-16
Activities and Deliverables	1-22
The Technology Landscape	1-26

Data Warehouse Automation Basics

Data Warehouse Automation Defined

Data Warehouse Automation (DWA) uses technology to gain efficiencies and improve effectiveness in data warehousing processes. DWA is much more than simply automating the development process. It encompasses all of the core processes of data warehousing including design, development, testing, deployment, operations, impact analysis, and change management.



- source data exploration
- warehouse data models
- schema generation
- ETL generation
- document generation
- metadata management
- test automation
- managed deployment
- scheduling & runtime operations
- change impact analysis
- maintenance & modification

Data Warehouse Automation Basics

Data Warehouse Automation Defined

COVERING THE WAREHOUSING LIFECYCLE

Data warehouse automation is more than simply automation of ETL development, or even the entire development process. It encompasses the entire data warehousing lifecycle from planning, analysis, and design through development and extending into operations, maintenance, and change management.

CHANGING BEST PRACTICES

Adoption of data warehouse automation changes the way that we think about building data warehouses. The widely accepted best practice of extensive up-front analysis, design, and modeling can be left behind as the mindset changes from “get it right the first time” to “develop fast and develop frequently” – an approach that is aligned with today’s agile development practices.

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

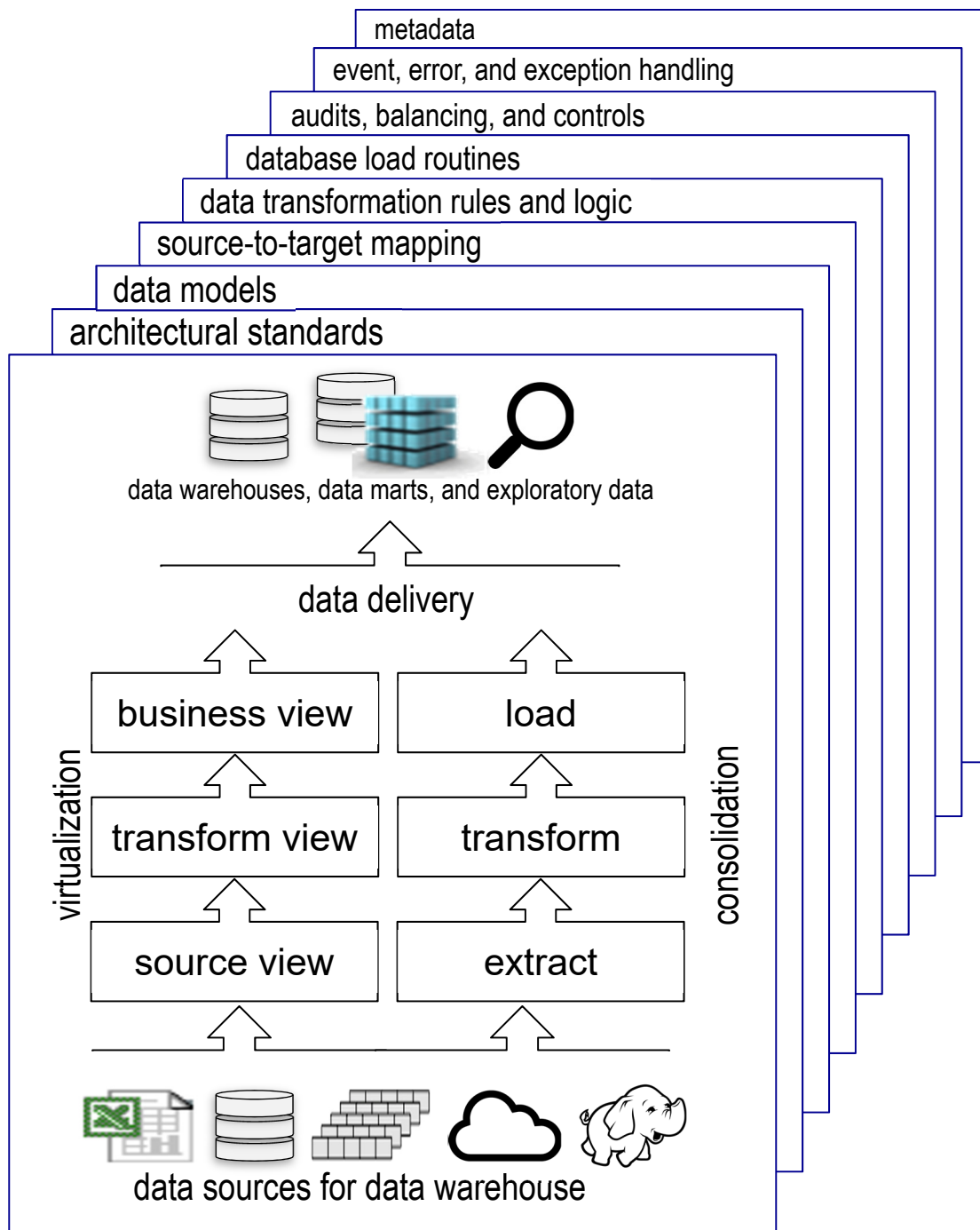
Speed is the critical factor that enables agility both for agile business and for agile development. The ability to generate quickly and to regenerate equally fast when change occurs is a fundamental automation capability.

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.

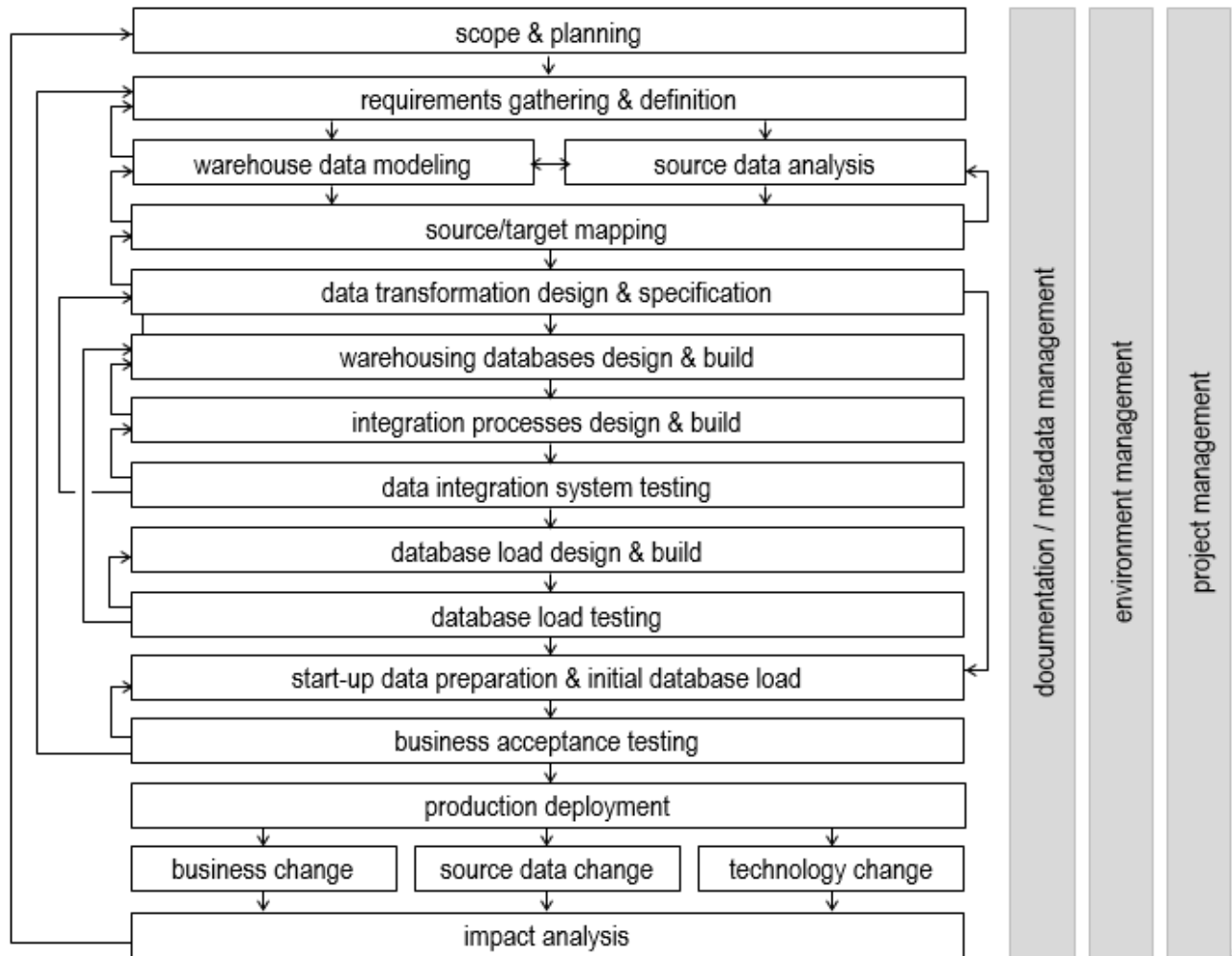
The center of the diagram shows the processing steps to get from data sources to integrated information using two methods – data consolidation through extract, transform, and load (ETL) processing and data virtualization through a series of abstract data views. Consolidation and virtualization are often 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.

Activities and Deliverables

Data Warehousing Processes



Activities and Deliverables

Data Warehousing Processes

LABOR INTENSITY, COMPLEXITY, AND VOLATILITY

A typical data warehouse development process is burdened by labor-intensive activities, a high degree of complexity and dependency among deliverables, and volatility of data and requirements. Some of the data warehousing pain that can be relieved by automation includes:

- Slow, laborious, and difficult to “get it right” requirement gathering
- Interdependency between source data analysis and warehouse data modeling
- Manually mapping sources to targets
- Detailed specification of data transformation logic
- Translating data models into schema and building databases
- Hand coding and manual testing of ETL processing
- Preparing and loading initial start-up data
- Acceptance testing and deployment to production

These activities would be difficult enough in a linear process, but in reality data warehousing projects are never linear. We are always going back because:

- Data modeling may change requirements
- Source-target mapping may affect data modeling
- Data transformation design may change mappings
- Integration process design may affect database design
- Incremental testing may cause changes throughout
- Acceptance testing may discover incorrect requirements
- And so on ...

When we survive all of these things and deploy the data warehouse it does not remain stable. Business needs, source data, and technology will all change (at different times and speeds), driving continuous change in the warehouse.

Through it all we need to manage projects, track changes, manage code and system versions, isolate production from development and testing environments, sustain daily operations, and try to keep metadata and documentation synchronized with what is actually implemented.

The Technology Landscape

Automation Tools and Vendors



WhereScape®

Data Warehouses. Right. Now.



The Technology Landscape

Automation Tools and Vendors

COMMON TOOLS

The facing page illustrates the four most common tools for data warehouse automation – Attunity Compose, Magnitude Software Kalido, TimeXtender, and WhereScape. These are the “pure” DWA product vendors. Others, such as Infor Birst and erwin Data Modeler, include some automation capabilities in a broader product suite.

Later in the course you will see more about the approaches and capabilities of Attunity Compose, Magnitude Software Kalido, TimeXtender, and WhereScape – each with different perspectives and capabilities. The intent is not to compare the tools or to perform product evaluation, but to show a representative sample of mainstream DWA technology.



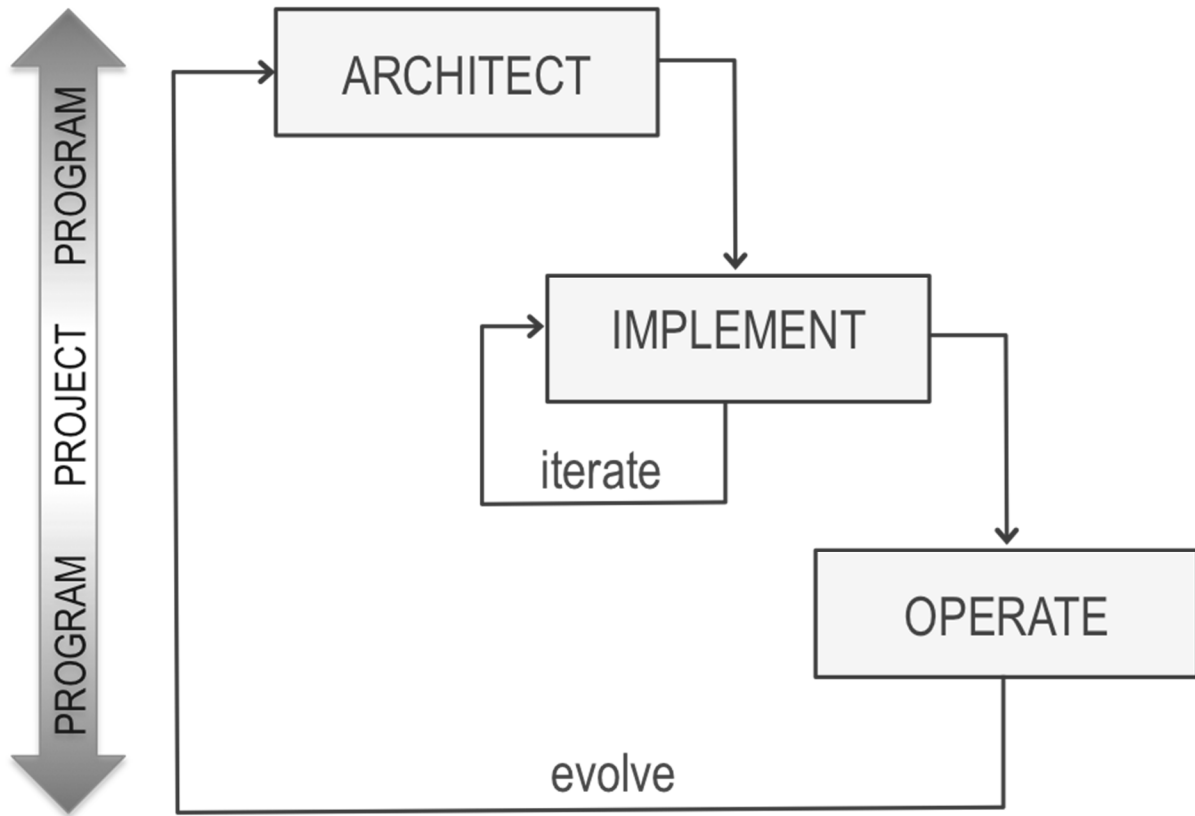
Module 2

Building and Managing the Data Warehouse

Topic	Page
The Data Warehousing Lifecycle	2-2
Building the Data Warehouse	2-22
Managing the Data Warehouse	2-30

The Data Warehousing Lifecycle

A Closed Loop



The Data Warehousing Lifecycle

A Closed Loop

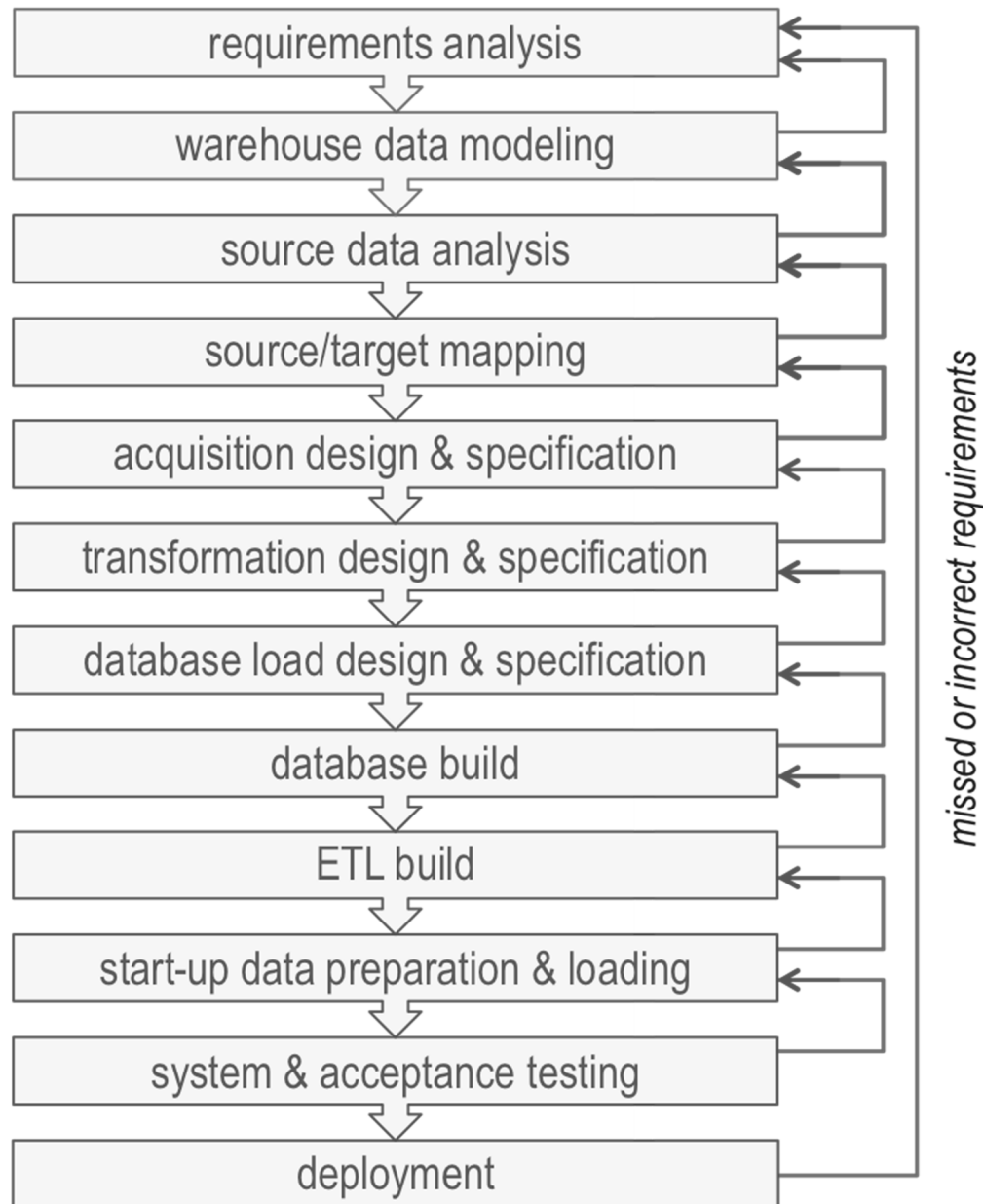
CONTINUOUS WITHOUT AN EXIT POINT

The diagram on the facing page revisits the data warehousing lifecycle with emphasis on the blending of program and project work. The important introductory concept is that the lifecycle is a closed-loop system. There is no exit point in the diagram and no end to continuous evolution, change, and redevelopment.

This is the fundamental reason that conventional design-and-build cycles do not work well for data warehousing.

Building the Data Warehouse

Requirements-Driven Development



Building the Data Warehouse

Requirements-Driven Development

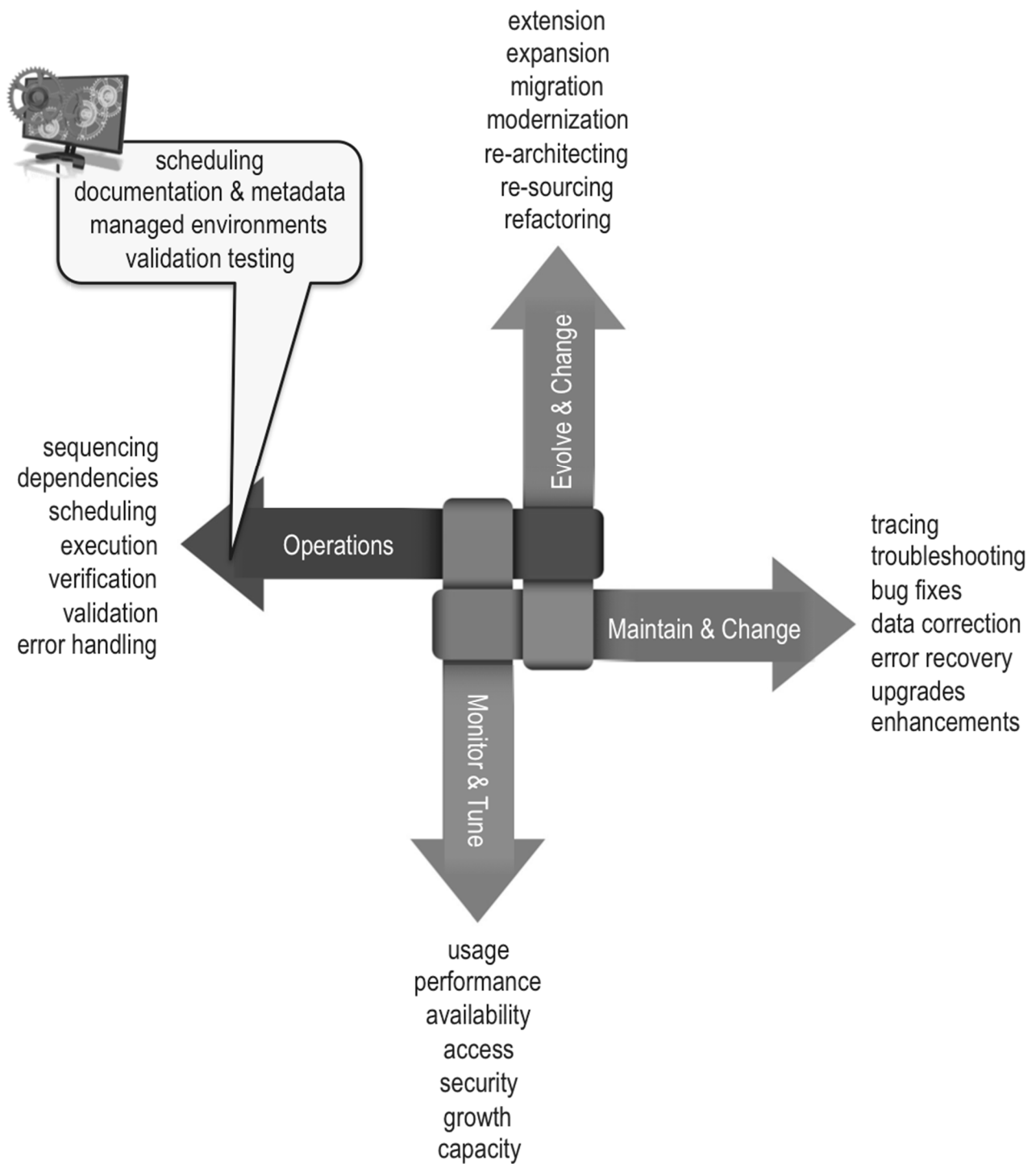
A TRADITIONAL APPROACH

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



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



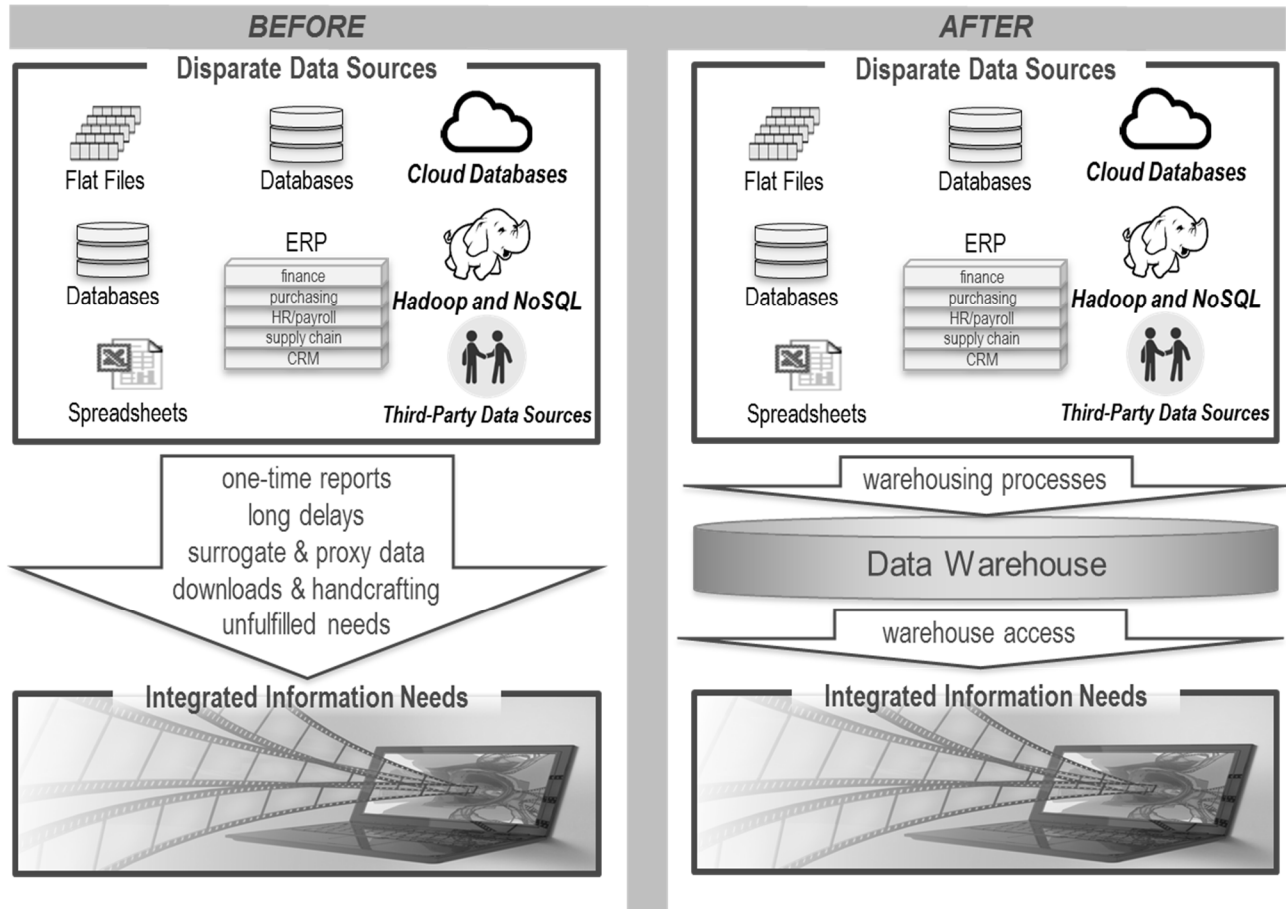
Module 3

Using Data Warehouse Automation

Topic	Page
Automation Use Cases	3-2
Case Studies	3-20

Automation Use Cases

Building a New Data Warehouse



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.

Regulatory reporting and audit demand quality data.

Requirement to manage both global and local brands.

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

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.

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

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

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

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

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

Topic	Page
Attunity Compose in Action	4-2
Magnitude Software Kalido in Action	4-4
TimeXtender in Action	4-6
WhereScape in Action	4-8



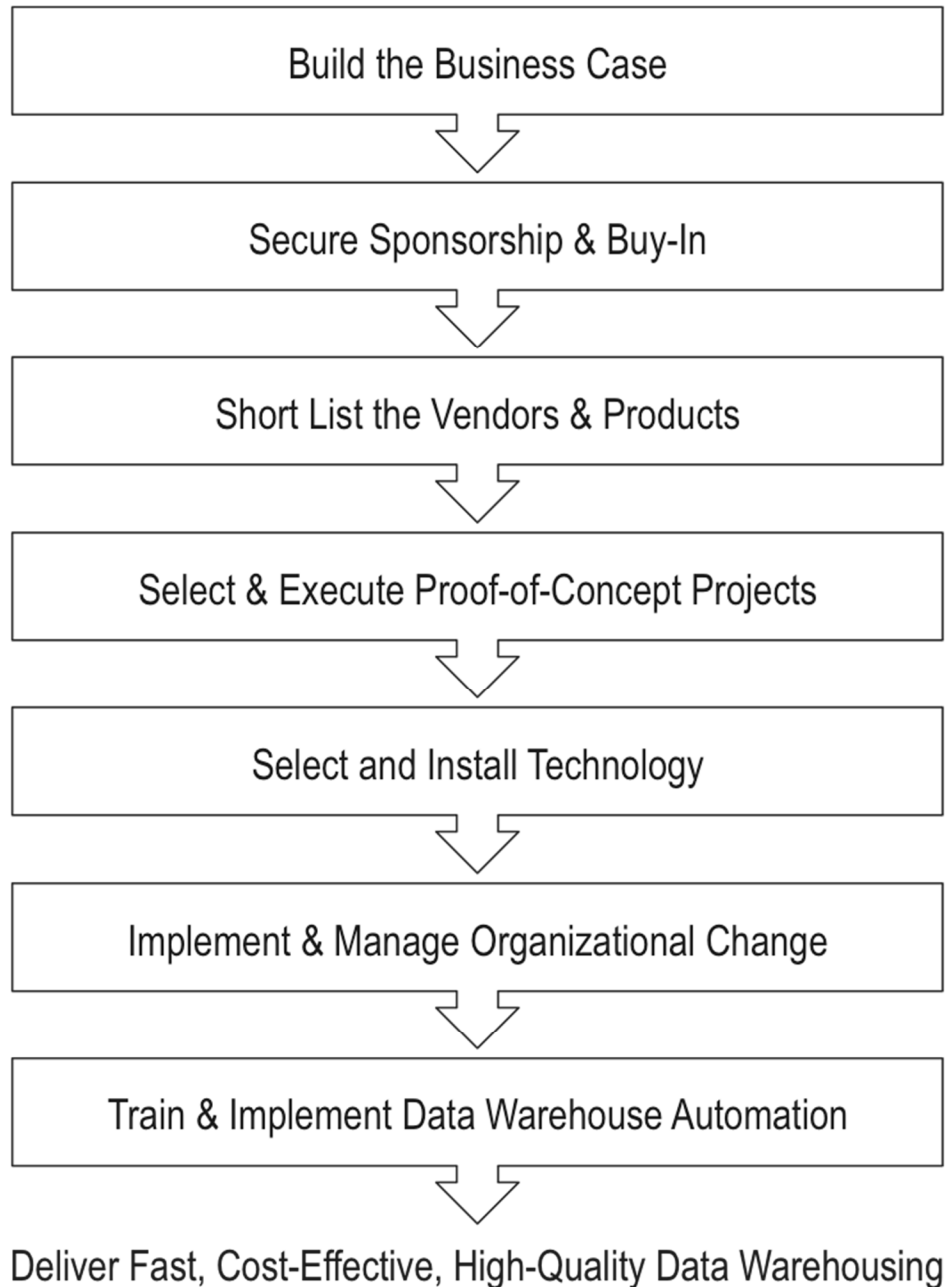
Module 5

Getting Started with Data Warehouse Automation

Topic	Page
Step by Step	5-2
Human Factors	5-4
New Horizons	5-6
Next Steps	5-8

Step by Step

A Process for Automation



Step by Step

A Process for Automation

BUSINESS CASE	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.
SPONSORSHIP AND BUY-IN	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.
SHORT LIST	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.
PROOF OF CONCEPT	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.
TECHNOLOGY SELECTION	Based on proof of concept results and previously defined selection criteria, make your technology decision and install the automation tools.
ORGANIZATIONAL CHANGE	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.
TRAINING AND IMPLEMENTATION	Train people to use the automation tools, and train them to understand their new roles and responsibilities with data warehouse automation.

