

Oracle Database 11g: Data Warehousing Fundamentals

Volume I • Student Guide

D56261GC10

Edition 1.0

February 2009

D58420

ORACLE®

Author

Lauran K. Serhal

**Technical Contributors
and Reviewers**

David Allan

Hermann Baer

Herbert Bradbury

Harald Van Breederode

Yanti Chang

Joel Goodman

Richard.Green

Nancy Greenberg

Martin Gubar

Yash Jain

Gerry Jurens

Dr. Thomas Merz

Brian Pottle

Jan Van Stappen

S Matt Taylor

Jean-Francois Verrier

Andreas Walter2

Editors

Nita Pavitran

Raj Kumar

Graphic Designer

Rajiv Chandrabhanu

Publishers

Sujatha Nagendra

Veena Narasimhan

Copyright © 2009, Oracle. All rights reserved.

Disclaimer

This document contains proprietary information and is protected by copyright and other intellectual property laws. You may copy and print this document solely for your own use in an Oracle training course. The document may not be modified or altered in any way. Except where your use constitutes "fair use" under copyright law, you may not use, share, download, upload, copy, print, display, perform, reproduce, publish, license, post, transmit, or distribute this document in whole or in part without the express authorization of Oracle.

The information contained in this document is subject to change without notice. If you find any problems in the document, please report them in writing to: Oracle University, 500 Oracle Parkway, Redwood Shores, California 94065 USA. This document is not warranted to be error-free.

Restricted Rights Notice

If this documentation is delivered to the United States Government or anyone using the documentation on behalf of the United States Government, the following notice is applicable:

U.S. GOVERNMENT RIGHTS

The U.S. Government's rights to use, modify, reproduce, release, perform, display, or disclose these training materials are restricted by the terms of the applicable Oracle license agreement and/or the applicable U.S. Government contract.

Trademark Notice

Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

Contents

Preface

1 Introduction

Objectives 1-2

Questions About You 1-3

Lesson Agenda 1-4

Course Objectives 1-5

Prerequisites and Suggested Prerequisites 1-7

Suggested Course Schedule 1-8

Lesson Agenda 1-10

Some of the Sample Schemas Used in the Course 1-11

Sales History (sh) Schema 1-12

Class Account Information 1-14

Appendixes in the Course 1-15

SQL Environments Available in the Course 1-16

Entering SQL Statements Using Oracle SQL*Plus 1-17

What Is Oracle SQL Developer? 1-18

Creating Schema Objects 1-19

Data Warehousing Tools Used in this Course 1-20

Lesson Agenda 1-21

Oracle 11g Useful Documentation 1-22

Continuing Your Education: Recommended Follow-Up Classes 1-24

Oracle Technology Network (OTN) 1-25

Oracle by Example (OBE): <http://www.oracle.com//technology/obe> 1-26

Oracle Warehouse Builder OBEs:

http://www.oracle.com/technology/obe/11gr1_owb/index.htm 1-27

Analytic Workspace Manager (AWM) OBEs:

<http://www.oracle.com/technology/products/bi/olap/olap.html> 1-28

Oracle Application Express (APEX) OBEs:

http://www.oracle.com/technology/obe/hol08/apex31nf/apexusr31/apexusr31_otn.htm 1-29

Technical Support: My Oracle Support: <https://metalink.oracle.com/CSP/ui/index.html> 1-30

Summary 1-31

Practice 1: Overview 1-32

2 Data Warehousing, Business Intelligence, OLAP, and Data Mining

Objectives	2-2
Lesson Agenda	2-3
Evolution of BI	2-4
Early Management Information Systems	2-5
Analyzing Data from Operational Systems	2-6
Why OLTP Is Not Suitable for Analytical Reporting	2-7
Data Extract Processing	2-8
Issues with Data Extract Programs	2-9
Productivity Issues with Extract Processing	2-10
Data Quality Issues with Extract Processing	2-11
Data Warehousing and Business Intelligence	2-12
Technological Solutions for Data Warehousing	2-13
Advantages of Warehouse Processing Environments	2-15
Business Intelligence (BI): Definition and Purpose	2-16
What Is Business Intelligence?	2-17
The User Dilemma	2-18
Problem: Multivendor, Unintegrated Environment	2-19
The Optimal Information Platform for Business Intelligence	2-20
The Optimal Information Platform	2-21
Success Factors for a Dynamic Business Environment	2-23
Business Drivers for Data Warehouses	2-24
Business Intelligence: Requirements	2-25
Quiz	2-26
Practice 2-1: Overview	2-27
Lesson Agenda	2-30
OLAP: Overview	2-31
Typical Example of an OLAP Query	2-32
The Dimensional Model	2-34
Measures	2-35
Measure Types	2-36
Dimensions	2-37
Example of Dimensions in a Report	2-38
Sharing Dimensions	2-39
Hierarchy	2-40
Hierarchy: Example	2-42
Level	2-43
Hierarchy Types	2-44
Attributes	2-45
Attributes and Levels: Examples	2-46
Dimensional Model Summarized	2-47

Physical Layers of a Data Warehouse 2-48
Managing the Data Warehouse 2-49
Oracle OLAP: Part of an Integrated DW Platform 2-50
What Is Oracle OLAP? 2-51
Oracle OLAP Option: 11g Highlights 2-52
ROLAP Versus MOLAP 2-53
Oracle Data Mining: An Overview 2-56
Oracle Data Mining: Interfaces 2-58
Quiz 2-59
Summary 2-60
Practice 2-2: Overview 2-61
Practice 2-3: Overview 2-63

3 Defining Data Warehouse Concepts and Terminology

Objectives 3-2
Lesson Agenda 3-3
Data Warehouse: Definition 3-4
Data Warehouse Properties 3-6
Subject Oriented 3-7
Integrated 3-8
Time Variant 3-10
Nonvolatile 3-11
Changing Warehouse Data 3-12
Data Warehouse Versus OLTP 3-13
Enterprise-Wide Data Warehouse 3-15
Data Warehouses Versus Data Marts 3-16
Dependent Data Mart 3-18
Independent Data Mart 3-19
Typical Data Warehouse Components 3-20
Quiz 3-22
Lesson Agenda 3-24
Warehouse Development Approaches 3-25
“Big Bang” Approach 3-26
Top-Down Approach 3-28
Bottom-Up Approach 3-29
Incremental Approach to Warehouse Development 3-31
Data Warehousing Process Components 3-32
Methodology 3-33
Architecture 3-34
Extraction, Transformation, and Loading (ETL) 3-35
Implementation 3-36

Operation and Support 3-37
Phases of the Incremental Approach 3-38
Strategy Phase Deliverables 3-40
Introducing the Case Study: Roy Independent School District (RISD) 3-42
Summary 3-44
Practice 3-1: Overview 3-45

4 Business, Logical, Dimensional, and Physical Modeling

Objectives 4-2
Lesson Agenda 4-3
Data Warehouse Modeling Issues 4-4
Data Warehouse: Design Phases 4-6
Phase 1: Defining the Business Model 4-7
Defining the Business Model: Performing Strategic Analysis 4-8
Defining the Business Model: Creating the Business Model 4-9
Business Requirements Drive the Design Process 4-10
Using a Business Process Matrix 4-11
Identifying Business Measures and Dimensions 4-12
Determining Granularity 4-14
Identifying Business Definitions and Rules: Example 4-15
Documenting Metadata 4-16
Business Metadata Elements 4-17
Metadata Documentation Approaches 4-18
Phase 2: Designing the Logical Model 4-19
Quiz 4-21
Practice 4-1: Overview 4-22
Lesson Agenda 4-24
Phase 3: Defining the Dimensional Model 4-25
Data Warehouse Schemas 4-26
Star Schema Model 4-27
Star Dimensional Modeling 4-28
Advantages of Using a Star Dimensional Model 4-29
Snowflake Schema Model 4-30
Third Normal Form (3NF) 4-32
Fact Table: Characteristics 4-33
More on Factless Fact Tables 4-35
Identifying Base and Derived Measures 4-36
Fact Table Measures 4-38
Dimension Table: Characteristics 4-39
Translating Business Dimensions into Dimension Tables 4-40
Slowly Changing Dimensions 4-41

Slowly Changing Dimension (SCD): An Example	4-42
Types of Database Keys	4-43
Using Time in the Data Warehouse	4-45
Time Dimension	4-46
Identifying Hierarchies for Dimensions	4-48
Using Hierarchies to Drill on Data and Aggregate Data	4-49
Using Data-Modeling Tools	4-50
Phase 4: Defining the Physical Model	4-52
Translating a Dimensional Model to a Physical Model	4-53
Architectural Requirements	4-55
Making the Right Choice	4-56
Storage and Performance Considerations	4-57
Quiz	4-58
Summary	4-59
Practice 4-2: Overview	4-60
Practice 4-3: Overview	4-64
Practice 4-4: Overview	4-65

5 Database Sizing, Storage, Performance, and Security Considerations

Objectives	5-2
Lesson Agenda	5-3
Sizing the Database and Other Storage Requirements	5-4
Estimating the Database Size	5-5
Validating Database Size Assumptions	5-6
Testing Load Sampling	5-7
Oracle Database Architectural Advantages	5-8
Lesson Agenda	5-9
Why Data Partitioning Is Needed	5-10
Data Partitioning	5-11
Benefits of Partitioning	5-13
Oracle's Partitioning Strategies	5-14
Partition Performance Benefits: Partition Pruning	5-16
Indexing	5-17
B-Tree Index	5-18
Bitmap Indexes	5-19
Bitmap Index: Example	5-20
Comparing B-Tree and Bitmap Indexes	5-22
Other Index Properties	5-23
Optimizing Star Queries: Tuning Star Queries	5-25
Optimizing Star Queries: Star Transformation	5-26
Star Query: Example	5-27

Quiz 5-28
Lesson Agenda 5-29
Parallelism 5-30
Degree of Parallelism (DOP) 5-31
Operations That Can Be Parallelized 5-32
Parallel Execution Server Pool 5-33
PARALLEL Clause: Examples 5-34
Using Summary Data 5-35
HP Oracle Exadata Storage Server Hardware 5-36
Exadata Storage Deployment 5-37
Exadata Key Benefits for Data Warehousing 5-39
Security in Data Warehouses 5-40
Oracle's Strategy for Data Warehouse Security 5-42
Oracle-Supplied Technology and Tools for Implementing VPD 5-45
Quiz 5-47
Summary 5-48
Practice 5-1: Overview 5-49

6 The ETL Process: Extracting Data

Objectives 6-2
Lesson Agenda 6-3
Extraction, Transformation, and Loading (ETL) Process 6-4
ETL: Tasks, Importance, and Cost 6-5
Extracting Data 6-7
Examining Data Sources 6-8
Production Data 6-9
Archive Data 6-10
Internal Data 6-11
External Data 6-12
Mapping Data 6-14
Lesson Agenda 6-15
Extraction Methods 6-16
Change Data Capture Mechanism in Oracle Database 6-18
Extraction Techniques 6-21
Designing Extraction Processes 6-22
Maintaining Extraction Metadata 6-24
Possible ETL Failures 6-25
Maintaining ETL Quality 6-26
Oracle's ETL Tool: Oracle Warehouse Builder 6-27
Oracle-Supported Features for ETL 6-28
Oracle's Solution for ETL: Oracle Streams, Replication, and Message Queuing 6-30

Quiz 6-32
Summary 6-33
Practice 6-1: Overview 6-34
Practice 6-2: Overview 6-36

7 The ETL Process: Transforming Data

Objectives 7-2
Lesson Agenda 7-3
Transformation 7-4
Remote Staging Model 7-5
On-Site Staging Model 7-6
Data Anomalies 7-7
Transformation Routines 7-8
Transforming Data: Problems and Solutions 7-9
Multipart Keys Problem 7-10
Multiple Local Standards Problem 7-11
Multiple Files Problem 7-12
Missing Values Problem 7-13
Duplicate Values Problem 7-14
Element Names Problem 7-15
Element Meanings Problem 7-16
Input Formats Problem 7-17
Referential Integrity Constraints Problem 7-18
Name and Address Problem 7-19
Name-and-Address Processing in Oracle Warehouse Builder 7-21
Quality Data: Importance and Benefits 7-23
Quality: Standards and Improvements 7-24
Data Quality Guidelines 7-26
Data Quality: Solutions and Management 7-27
Lesson Agenda 7-28
Transformation Techniques: Merging Data 7-29
Transformation Techniques: Adding a Date Stamp 7-30
Transformation Techniques: Adding Keys to Data 7-31
Summarizing Data 7-32
Maintaining Transformation Metadata 7-33
Data Ownership and Responsibilities 7-35
Transformation Timing and Location 7-36
Choosing a Transformation Point 7-37
Monitoring and Tracking 7-38
Designing Transformation Processes 7-39
Transformation Tools 7-40

- Data Transformation 7-41
- Multistage Data Transformation 7-42
- Pipelined Data Transformation 7-43
- Oracle's Enhanced Features for Transformation 7-44
- Application of the MERGE Statement in Data Warehousing: Example 7-45
- Multitable INSERT Statements 7-46
- Advantages of Multitable INSERTs 7-47
- Oracle's Enhanced Features for Transformation 7-48
- Summary 7-49
- Practice 7-1: Overview 7-50

8 The ETL Process: Loading Data

- Objectives 8-2
- Lesson Agenda 8-3
- Loading Data into the Warehouse 8-4
- Transportation in a Data Warehouse 8-6
- Transportable Tablespaces 8-7
- Initial Load and Refresh 8-8
- Data Refresh Models: Extract Processing Environment 8-10
- Data Refresh Models: Warehouse Processing Environment 8-11
- Building the Loading Process 8-12
- Data Granularity 8-13
- Loading Techniques 8-14
- Loading Technique Considerations 8-15
- Loading Techniques Provided by Oracle: SQL*Loader 8-17
- Loading Techniques Provided by Oracle 8-19
- Using External Tables 8-21
- Benefits of Using External Tables 8-22
- Creating External Tables 8-23
- Example of Defining External Tables 8-24
- Defining External Tables Using SQL*Loader 8-25
- Creating and Loading an External Table
 - Using ORACLE_DATAPUMP: Example 8-26
- Lesson Agenda 8-27
- Postprocessing of Loaded Data 8-28
- Indexing and Sorting Data 8-29
- Unique Indexes 8-30
- Creating Derived Keys 8-31
- Summary Management 8-33
- Filtering Data 8-34

Verifying Data Integrity 8-35
Steps for Verifying Data Integrity 8-36
Standard Quality Assurance Checks 8-38
Quiz 8-39
Summary 8-41
Practice 8-1: Overview 8-42

9 Refreshing the Warehouse Data

Objectives 9-2
Lesson Agenda 9-3
Developing a Refresh Strategy for Capturing Changed Data 9-4
User Requirements and Assistance 9-5
Load Window Requirements 9-6
Planning the Load Window 9-7
Scheduling the Load Window 9-8
Capturing Changed Data for Refresh 9-12
Choosing a Method for Change Data Capture 9-14
Wholesale Data Replacement 9-15
Comparison of Database Instances 9-16
Time- and Date-Stamping 9-17
Database Triggers 9-18
Using a Database Log 9-19
Refresh Mechanisms in the Oracle Database 9-20
Lesson Agenda 9-22
Applying the Changes to Data 9-23
Overwriting a Record 9-24
Adding a New Record 9-25
Adding a Current Field 9-26
Limitations of Methods for Applying Changes 9-27
Maintaining History: Techniques 9-29
History Tables and One-to-Many Relationships 9-30
Versioning 9-32
Preserving Complete History 9-33
Purging and Archiving Data 9-34
Oracle-Supported Techniques for Purging Data 9-35
Oracle-Supported Techniques for Archiving Data 9-36
Final Tasks 9-37
Publishing Data 9-38
Quiz 9-39
Summary 9-40
Practice 9-1: Overview 9-41

10 Materialized Views

- Objectives 10-2
- Lesson Agenda 10-3
- The Need for Summary Management 10-4
- Using Summaries to Improve Performance 10-5
- Summary Management 10-6
- Summary Navigation 10-7
- Managing Historical Summary Data in the Warehouse 10-8
- Summary Management Using the Oracle Database 10-9
- Using Materialized Views for Summary Management 10-11
- Using Summaries Without Materialized Views: Example 10-12
- Using Materialized Views for Summary Management: Example 10-13
- Determining Which Materialized View to Create 10-14
- Creating a Materialized View Using the `CREATE SQL` Statement: Example 10-15
- Available Refresh Modes When Creating a Materialized View 10-16
- Manual Refresh Using the `DBMS_MVIEW` Package Procedures 10-17
- Using the `DBMS_MVIEW` Package: The Available `ON DEMAND` Refresh Methods 10-18
- Refreshing at Scheduled Time: Using the `START WITH` and `NEXT` Clauses 10-20
- Query Rewrite Overview 10-21
- Cost-Based Query Rewrite Process 10-22
- Conditions Required for Oracle to Rewrite a Query 10-23
- Query Rewrite 10-24
- Quiz 10-25
- Lesson Agenda 10-26
- What Are Dimensions? 10-27
- What Are Dimension Objects? 10-28
- Why Are Dimensions Important? 10-29
- Dimensions and Hierarchies 10-30
- Dimension Example 10-31
- Defining Dimensions and Hierarchies 10-32
- Dimensions with Multiple Hierarchies 10-33
- Quiz 10-34
- Summary 10-35
- Practice 10-1: Overview 10-36

11 Leaving a Metadata Trail

- Objectives 11-2
- Defining Warehouse Metadata 11-3

Metadata Users	11-5
Types of Metadata	11-6
Examining Metadata: ETL Metadata	11-7
Extraction Metadata	11-8
Transformation Metadata	11-9
Loading Metadata	11-10
Examining Metadata: End-User Metadata	11-11
End-User Metadata: Context	11-12
Historic Context of Data	11-13
Types of Context	11-14
Developing a Metadata Strategy	11-15
Defining Metadata Goals and Intended Usage	11-16
Identifying Target Metadata Users	11-17
Choosing Metadata Tools and Techniques	11-18
Choosing the Metadata Location	11-20
Managing the Metadata	11-21
Integrating Multiple Sets of Metadata	11-22
Managing Changes to Metadata	11-23
Additional Metadata Content and Considerations	11-24
Common Warehouse Metamodel	11-26
Oracle Warehouse Builder: Compliance with OMG-CWM	11-27
Quiz	11-29
Summary	11-30
Practice 11-1: Overview	11-31

12 Data Warehouse Implementation Considerations

Objectives	12-2
Project Management	12-3
Requirements Specification or Definition	12-5
Logical, Dimensional, and Physical Data Models	12-6
Data Warehouse Architecture	12-8
ETL Considerations	12-9
Reporting Considerations	12-10
Security Considerations	12-11
Metadata Management	12-12
Testing the Implementation	12-14
Post-Implementation Change Management	12-15
Some Useful Resources and White Papers	12-17
Summary	12-18

Appendix A: Practices and Solutions

Appendix B: Table Descriptions

Appendix C: Using SQL Developer

Objectives	C-2
What Is Oracle SQL Developer?	C-3
Specifications of SQL Developer	C-4
Installing SQL Developer	C-5
SQL Developer Interface	C-6
Creating a Database Connection	C-7
Browsing Database Objects	C-10
Creating a Schema Object	C-11
Creating a New Table: Example	C-12
Using the SQL Worksheet	C-13
Executing SQL Statements	C-16
Saving SQL Scripts	C-17
Executing Saved Script Files: Method 1	C-18
Executing Saved Script Files: Method 2	C-19
Executing SQL Statements	C-20
Formatting the SQL Code	C-21
Using Snippets	C-22
Using Snippets: Example	C-23
Using SQL*Plus	C-24
Debugging Procedures and Functions	C-25
Database Reporting	C-26
Creating a User-Defined Report	C-27
Search Engines and External Tools	C-28
Setting Preferences	C-29
Summary	C-30

Appendix D: Using SQL*Plus

Objectives	D-2
SQL and SQL*Plus Interaction	D-3
SQL Statements Versus SQL*Plus Commands	D-4
Overview of SQL*Plus	D-5
Logging In to SQL*Plus: Available Methods	D-6
Customizing the SQL*Plus Environment	D-7
Displaying Table Structure	D-8
SQL*Plus Editing Commands	D-10
Using LIST, n, and APPEND	D-12

Using the <code>CHANGE</code> Command	D-13
SQL*Plus File Commands	D-14
Using the <code>SAVE</code> , <code>START</code> , and <code>EDIT</code> Commands	D-15
SQL*Plus Enhancements Since Oracle Database 10g	D-17
Changes to the <code>SERVEROUTPUT</code> Command	D-18
White Space Support in File and Path Names in Windows	D-19
Predefined SQL*Plus Variables	D-20
Using the New Predefined SQL*Plus Variables: Examples	D-21
<code>SHOW</code> Command and the New <code>RECYCLEBIN</code> Clause	D-22
<code>SHOW</code> Command and the New <code>RECYCLEBIN</code> Clause: Example	D-23
Using the SQL*Plus <code>SPOOL</code> Command	D-24
Using the SQL*Plus <code>SPOOL</code> Command: Examples	D-25
<code>COPY</code> Command: New Error Messages	D-26
Change in the <code>DESCRIBE</code> Command Behavior	D-29
<code>SET PAGES [IZE]</code> Command	D-30
<code>SQLPLUS</code> Program and the Compatibility Option	D-31
Using the <code>AUTOTRACE</code> Command	D-32
Displaying a Plan Table Using the <code>DBMS_XPLAN.DISPLAY</code> Package Function	D-33
Summary	D-34

Appendix E: Basic Linux and vi Commands

Index

Case Study

