

Oracle® Data Mining

Administrator's Guide

11g Release 1 (11.1)

B28130-02

September 2007

Oracle Data Mining Administrator's Guide, 11g Release 1 (11.1)

B28130-02

Copyright © 2005, 2007, Oracle. All rights reserved.

The Programs (which include both the software and documentation) contain proprietary information; they are provided under a license agreement containing restrictions on use and disclosure and are also protected by copyright, patent, and other intellectual and industrial property laws. Reverse engineering, disassembly, or decompilation of the Programs, except to the extent required to obtain interoperability with other independently created software or as specified by law, is prohibited.

The information contained in this document is subject to change without notice. If you find any problems in the documentation, please report them to us in writing. This document is not warranted to be error-free. Except as may be expressly permitted in your license agreement for these Programs, no part of these Programs may be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose.

If the Programs are delivered to the United States Government or anyone licensing or using the Programs on behalf of the United States Government, the following notice is applicable:

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the Programs, including documentation and technical data, shall be subject to the licensing restrictions set forth in the applicable Oracle license agreement, and, to the extent applicable, the additional rights set forth in FAR 52.227-19, Commercial Computer Software--Restricted Rights (June 1987). Oracle USA, Inc., 500 Oracle Parkway, Redwood City, CA 94065.

The Programs are not intended for use in any nuclear, aviation, mass transit, medical, or other inherently dangerous applications. It shall be the licensee's responsibility to take all appropriate fail-safe, backup, redundancy and other measures to ensure the safe use of such applications if the Programs are used for such purposes, and we disclaim liability for any damages caused by such use of the Programs.

Oracle, JD Edwards, PeopleSoft, and Siebel are registered trademarks of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

The Programs may provide links to Web sites and access to content, products, and services from third parties. Oracle is not responsible for the availability of, or any content provided on, third-party Web sites. You bear all risks associated with the use of such content. If you choose to purchase any products or services from a third party, the relationship is directly between you and the third party. Oracle is not responsible for: (a) the quality of third-party products or services; or (b) fulfilling any of the terms of the agreement with the third party, including delivery of products or services and warranty obligations related to purchased products or services. Oracle is not responsible for any loss or damage of any sort that you may incur from dealing with any third party.

Contents

Preface	vii
Audience	vii
Documentation Accessibility	vii
Related Documentation	viii
Where to Find Database Tuning Documentation	viii
Conventions	ix
What's New in Oracle Data Mining Administration	xi
No DMSYS Schema	xi
Mining Models in the Oracle Data Dictionary	xi
Enhanced Security	xii
Scoping of Nested Data	xii
Enhanced Handling of Sparse Data	xii
Desupported and Deprecated Features	xii
1 Getting Started	
Install Oracle Database	1-1
Install Oracle Database Companion	1-7
Create a Data Mining Demo User	1-10
Run the Sample Programs	1-11
2 Connecting to a Database for Data Mining	
Your User ID	2-1
Connect to a Local Database	2-1
Connect to a Remote Database	2-2
Create a Net Service Name	2-3
3 Upgrading Oracle Data Mining	
Upgrading Oracle Data Mining	3-1
Downgrading Oracle Data Mining	3-3
4 Users and Privileges for Data Mining	
Creating Data Mining Users	4-1
System Privileges for Mining Models	4-2

Object Privileges on Mining Models	4-3
5 Installing the Spreadsheet Add-In for Predictive Analytics	
About the Spreadsheet Add-In.....	5-1
Requirements	5-1
Install the Spreadsheet Add-In.....	5-2
6 Mining Model Schema Objects	
Obtaining Information from the Data Dictionary	6-1
Adding a Comment to a Mining Model.....	6-3
Auditing Mining Models.....	6-4
Exporting and Importing Mining Models.....	6-5
7 The Data Mining Sample Programs	
PL/SQL Programs	7-1
Java Programs	7-3
Text Mining Programs	7-4
The Sample Data	7-7
Index	

Preface

This manual explains how to install the various components of Oracle Data Mining and perform basic administration tasks. It also explains how to install and run the Data Mining sample programs.

The preface contains these topics:

- [Audience](#)
- [Documentation Accessibility](#)
- [Related Documentation](#)
- [Where to Find Database Tuning Documentation](#)
- [Conventions](#)

Audience

This guide can be used by a spectrum of users, however, it is primarily directed at an individual user who wants to install, configure, and use Oracle Data Mining on a laptop or personal computer.

Documentation Accessibility

Our goal is to make Oracle products, services, and supporting documentation accessible, with good usability, to the disabled community. To that end, our documentation includes features that make information available to users of assistive technology. This documentation is available in HTML format, and contains markup to facilitate access by the disabled community. Accessibility standards will continue to evolve over time, and Oracle is actively engaged with other market-leading technology vendors to address technical obstacles so that our documentation can be accessible to all of our customers. For more information, visit the Oracle Accessibility Program Web site at

<http://www.oracle.com/accessibility/>

Accessibility of Code Examples in Documentation

Screen readers may not always correctly read the code examples in this document. The conventions for writing code require that closing braces should appear on an otherwise empty line; however, some screen readers may not always read a line of text that consists solely of a bracket or brace.

Accessibility of Links to External Web Sites in Documentation

This documentation may contain links to Web sites of other companies or organizations that Oracle does not own or control. Oracle neither evaluates nor makes any representations regarding the accessibility of these Web sites.

TTY Access to Oracle Support Services

Oracle provides dedicated Text Telephone (TTY) access to Oracle Support Services within the United States of America 24 hours a day, 7 days a week. For TTY support, call 800.446.2398. Outside the United States, call +1.407.458.2479.

Related Documentation

The documentation set for Oracle Data Mining is part of the Oracle Database 11g Release 1 (11.1) Online Documentation Library. The Oracle Data Mining documentation set consists of the following documents:

- *Oracle Data Mining Concepts*
- *Oracle Data Mining Application Developer's Guide*
- *Oracle Data Mining Java API Reference* (javadoc)

For detailed information about the Oracle Data Mining PL/SQL interface, see *Oracle Database PL/SQL Packages and Types Reference*. Search for DBMS_DATA_MINING.

For detailed information about the SQL data mining functions, see *Oracle Database SQL Language Reference*.

For an introduction to application development in SQL and PL/SQL, see *Oracle Database Advanced Application Developer's Guide*.

For an introduction to application development in Java, see *Oracle Database Java Developer's Guide*

Where to Find Database Tuning Documentation

The Oracle Data Mining documentation set does not address the topic of database tuning for data mining. The *Oracle Database Performance Tuning Guide* is the best source of information for tuning Oracle Database (including when the goal is to invoke mining build/score operations).

The memory tuning parameters PGA_AGGREGATE_TARGET and SGA_TARGET can easily have the most impact on data mining. The correct setting of PGA is very important for model building routines, as well as complex queries and batch scoring. From a Data Mining perspective, the SGA is generally less of a concern, except that real-time scoring is enabled by having the model loaded into the shared cursor in the SGA, so this still should be sized accordingly.

See Also:

Oracle Database Performance Tuning Guide for getting started

Oracle Database Performance Tuning Guide for more detail on memory configuration

In addition, if parallel execution is desired, then there are INIT.ORA parameters that control this functionality.

See Also: Chapter 25 in *Oracle Database Data Warehousing Guide* for information about parallel execution tuning

Conventions

The following text conventions are used in this document:

Convention	Meaning
boldface	Boldface type indicates graphical user interface elements associated with an action, or terms defined in text or the glossary.
<i>italic</i>	Italic type indicates book titles, emphasis, or placeholder variables for which you supply particular values.
monospace	Monospace type indicates commands within a paragraph, URLs, code in examples, text that appears on the screen, or text that you enter.

What's New in Oracle Data Mining Administration

This section summarizes the new features of Oracle Data Mining that pertain to installation, administration, and upgrade.

See Also:

- "What's New" in *Oracle Data Mining Concepts* for a complete list of new and deprecated features in Oracle Data Mining 11g Release 1 (11.1).
- [Chapter 3](#) for information on upgrading and a downgrading a database

This section contains the following topics:

- [No DMSYS Schema](#)
- [Mining Models in the Oracle Data Dictionary](#)
- [Enhanced Security](#)
- [Scoping of Nested Data](#)
- [Enhanced Handling of Sparse Data](#)
- [Desupported and Deprecated Features](#)

No DMSYS Schema

Oracle Data Mining 11g Release 1 (11.1) has a tight integration with Oracle Database. Data Mining metadata and PL/SQL packages have been migrated from DMSYS to SYS. The DMSYS schema no longer exists in Oracle Database 11g Release 1 (11.1) fresh installations.

Mining Models in the Oracle Data Dictionary

New catalog views for Data Mining are introduced in 11g Release 1 (11.1):

- ALL/DBA/USER_MINING_MODELS provides information about mining models
- ALL/DBA/USER_MINING_MODEL_ATTRIBUTES provides information about data columns used to build a mining model
- ALL/DBA/USER_MINING_MODEL_SETTINGS provides information about configuration settings for mining models

The ALL/DBA/USER_OBJECTS catalog view now identifies mining models.

See Also: ["Obtaining Information from the Data Dictionary"](#) in [Chapter 6](#).

Enhanced Security

Security features of Oracle Data Mining are significantly enhanced in 11g Release 1 (11.1). Improved security for data mining has several aspects:

- Previously, Oracle Data Mining used DMSYS as its own repository schema. This necessitated the granting of advanced database privileges to DMSYS, a non-system account. In 11g Release 1 (11.1), the Oracle Data Mining metadata repository is in SYS, where it is accessible only to the system DBA.
- New system and object privileges for mining model objects are introduced in 11g Release 1 (11.1).
- The SQL auditing system can be used to track operations on mining model objects.

Note: The privilege CREATE MINING MODEL is required for creating models in 11g. This privilege should be added to any accounts being upgraded to 11g.

See Also: [Chapter 4, "Users and Privileges for Data Mining"](#).

Scoping of Nested Data

Oracle Data Mining supports nested data types for both categorical and numerical data. Multi-record case data must be transformed to nested columns for mining.

In Oracle Data Mining 10gR2, nested columns were processed as top-level attributes; the user had to ensure that two nested columns did not contain an attribute with the same name. In Oracle Data Mining 11g, nested attributes are scoped with the column name, which relieves the user of this burden.

See Also: *Oracle Data Mining Application Developer's Guide*

Enhanced Handling of Sparse Data

Handling of sparse data and missing values has been standardized across algorithms in Oracle Data Mining 11g. Data is sparse when a high percentage of the cells are empty, but all the values are assumed to be known. Only nested data can be considered sparse. Missing values in simple numeric or character columns are considered missing at random.

See Also: *Oracle Data Mining Application Developer's Guide*

Desupported and Deprecated Features

The following features are desupported in 11g Release 1 (11.1):

- DMSYS schema
- Oracle Data Mining Scoring Engine

- In Oracle 10.2, you could use Database Configuration Assistant (DBCA) to configure the Data Mining option. In Oracle 11g, you do not need to use DBCA to configure the Data Mining option.
- Basic Local Alignment Search Tool (BLAST)

The following features are deprecated in 11g Release 1 (11.1):

- Adaptive Bayes Network classification algorithm
- DM_USER_MODELS view is replaced by data dictionary views
- Several PL/SQL procedures have been deprecated.

- GET_DEFAULT_SETTINGS

Replaced with data dictionary views:

USER/ALL/DBA_MINING_MODEL_SETTINGS

- GET_MODEL_SETTINGS

Replaced with data dictionary views:

USER/ALL/DBA_MINING_MODEL_SETTINGS

- GET_MODEL_SIGNATURE

Replaced with data dictionary views:

USER/ALL/DBA_MINING_MODEL_ATTRIBUTES

See Also: *Oracle Database PL/SQL Packages and Types Reference*

Note: Oracle recommends that you do not use deprecated procedures in new applications. Support for deprecated features is for backward compatibility only.

Getting Started

This chapter is intended for anyone who wishes to install and use Oracle Data Mining on a personal computer for educational or demonstration purposes. It provides the basic information you will need to install the Data Mining software on Microsoft Windows and run the sample programs locally on your pc or laptop. To run the programs remotely, see the instructions in [Chapter 2](#).

See Also:

Oracle Database 2 Day DBA

Oracle Database Installation Guide for Microsoft Windows

This chapter contains the following sections. Complete the instructions in each section before proceeding to the next section.

- [Install Oracle Database](#)
- [Install Oracle Database Companion](#)
- [Create a Data Mining Demo User](#)
- [Run the Sample Programs](#)

Install Oracle Database

Oracle Data Mining is part of Oracle Database Enterprise Edition. To perform data mining activities, you must be able to log on to an Oracle database, and your user ID must have the appropriate database privileges.

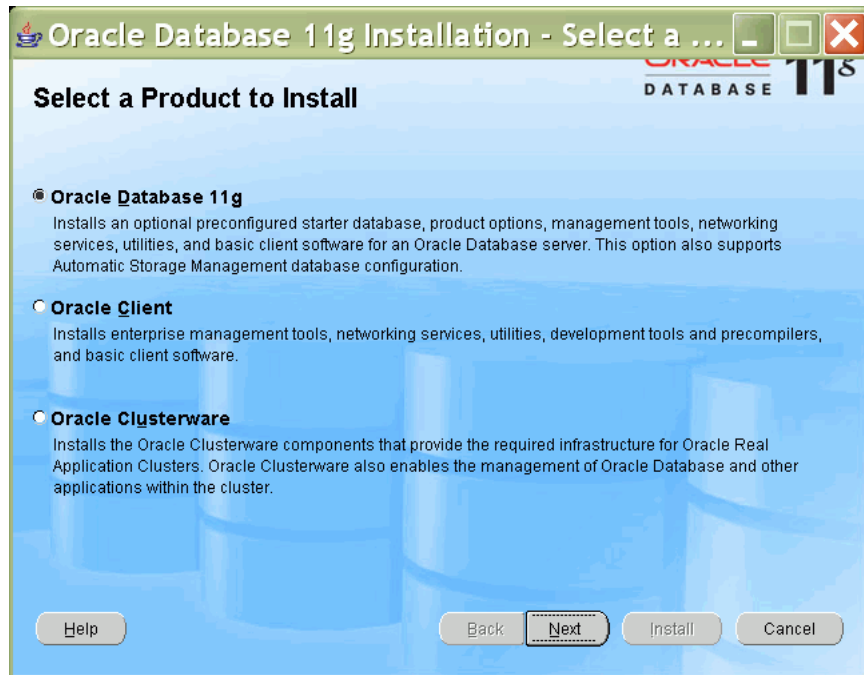
The instructions in this section explain how to install Oracle Database with the sample schemas on your personal computer. The sample schemas are needed for the Data Mining sample programs.

Note: These instructions assume that this is a fresh installation of Oracle Database 11g.

If you already have Oracle components installed on your computer, refer to *Oracle Database Installation Guide for Microsoft Windows*.

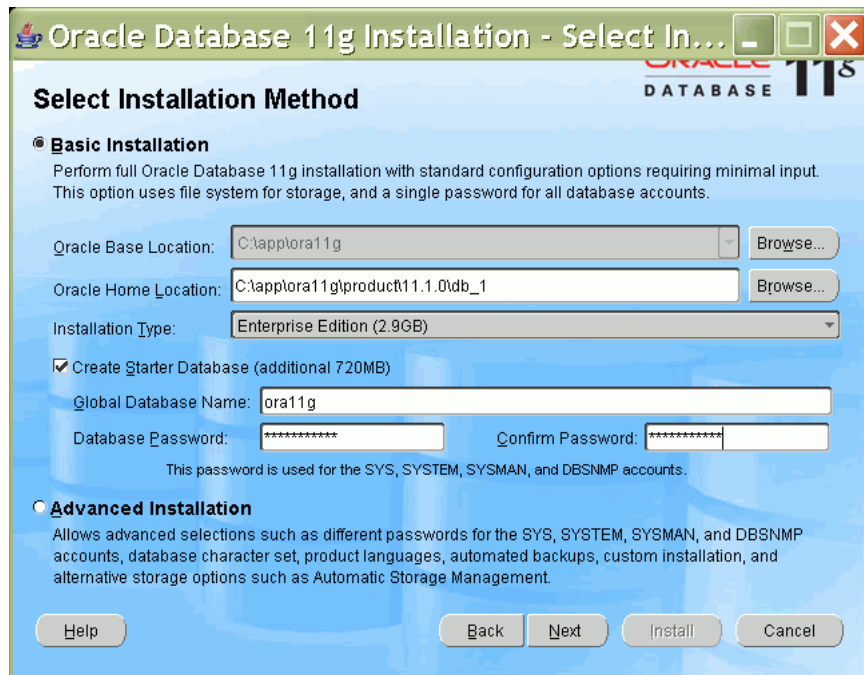
1. From the Database installation directory, run `SETUP . EXE`.

Oracle Universal Installer opens and displays the Select a Product to Install dialog. Choose **Oracle Database 11g**.



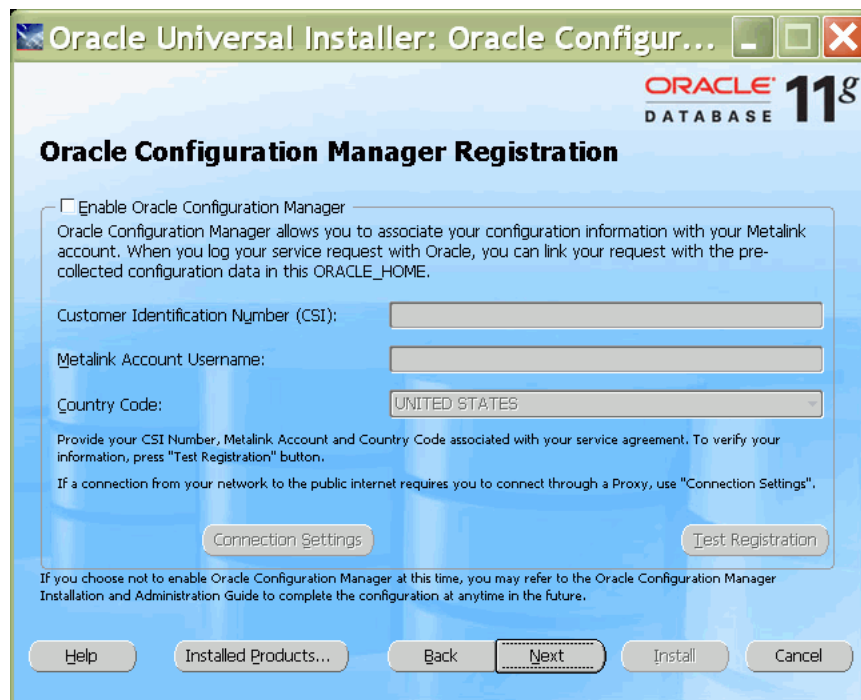
Choose **Next**.

- The Installer displays the Select Installation Method page.



- Choose **Basic Installation**.
- Specify the Oracle Base and Home directories. Oracle home is a subdirectory of the Oracle Base directory. You can accept the default paths provided by the Installer, as long as they do not already exist on your computer.
- Choose **Enterprise Edition** as the Installation Type.
- Check the **Create Starter Database** box.

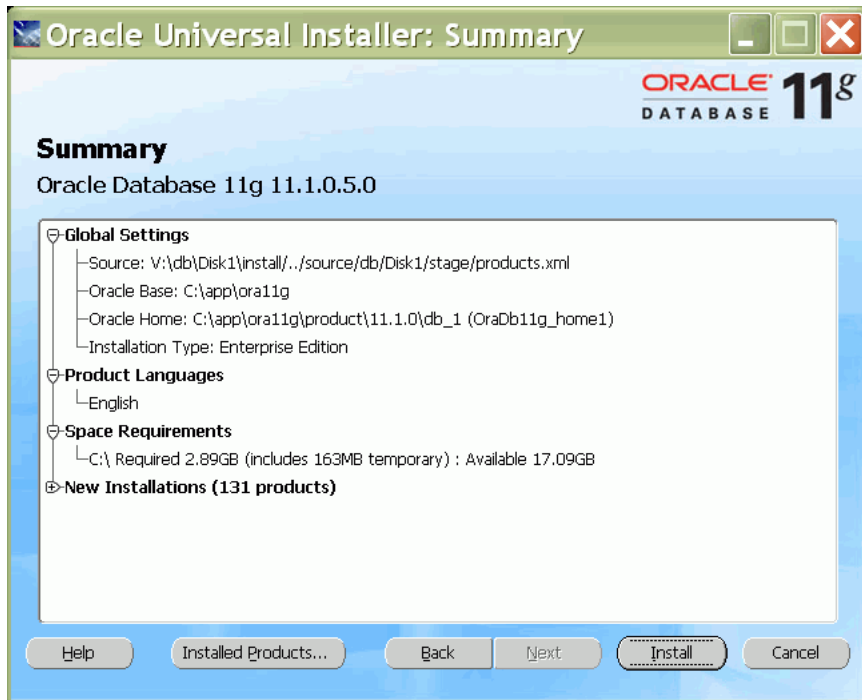
- Specify a unique name for **Global Database Name**. You can use the default global database name provided by the Installer, as long as it does not already exist on your computer.
 - Specify a password for the database accounts. The password must have at least eight characters and include both alphabetic and numeric characters. For details about specifying passwords, refer to *Oracle Database Security Guide*.
You will have the opportunity to change the passwords for the database accounts at a later time.
 - Click **Next**.
3. On the Oracle Configuration Manager Registration page, you can choose to register your installation with your Metalink account.



The screenshot shows the 'Oracle Configuration Manager Registration' dialog box from the Oracle Universal Installer. The title bar reads 'Oracle Universal Installer: Oracle Configur...'. The Oracle logo and 'ORACLE DATABASE 11g' are in the top right. The main heading is 'Oracle Configuration Manager Registration'. There is a checkbox labeled 'Enable Oracle Configuration Manager' which is currently unchecked. Below this, a text box explains that Oracle Configuration Manager allows associating configuration information with a Metalink account. There are three input fields: 'Customer Identification Number (CSI):', 'Metalink Account Username:', and 'Country Code:' (with a dropdown menu showing 'UNITED STATES'). A 'Test Registration' button is to the right of the Country Code field. Below these fields, there is a 'Connection Settings' button. At the bottom of the dialog, there are several navigation buttons: 'Help', 'Installed Products...', 'Back', 'Next' (which is highlighted with a dashed border), 'Install', and 'Cancel'. A small note at the bottom states: 'If you choose not to enable Oracle Configuration Manager at this time, you may refer to the Oracle Configuration Manager Installation and Administration Guide to complete the configuration at anytime in the future.'

This page is optional. You can simply choose **Next**.

4. The Summary page displays the settings and components for the installation.

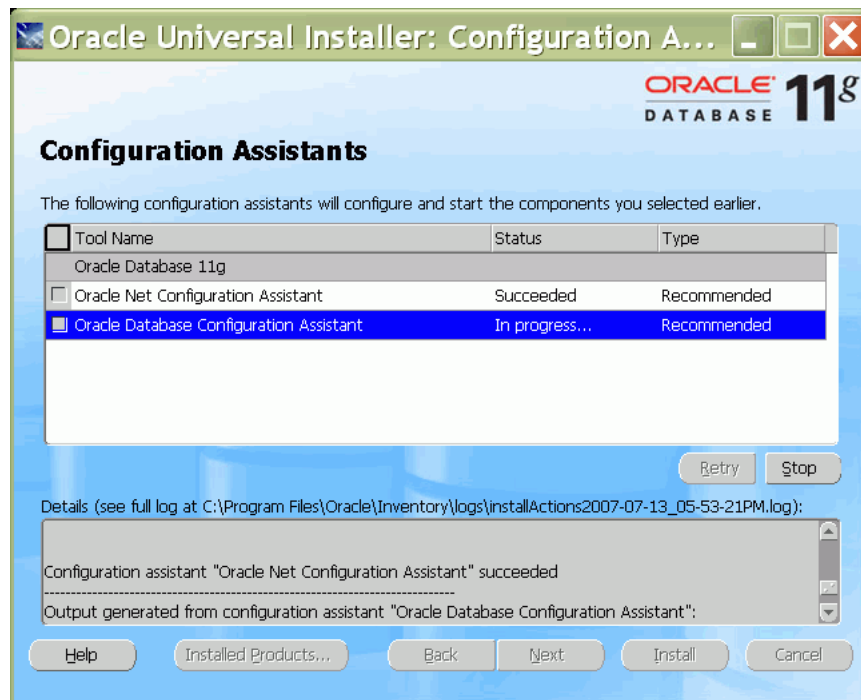


Click **Install**.

5. The Installer proceeds with the installation.

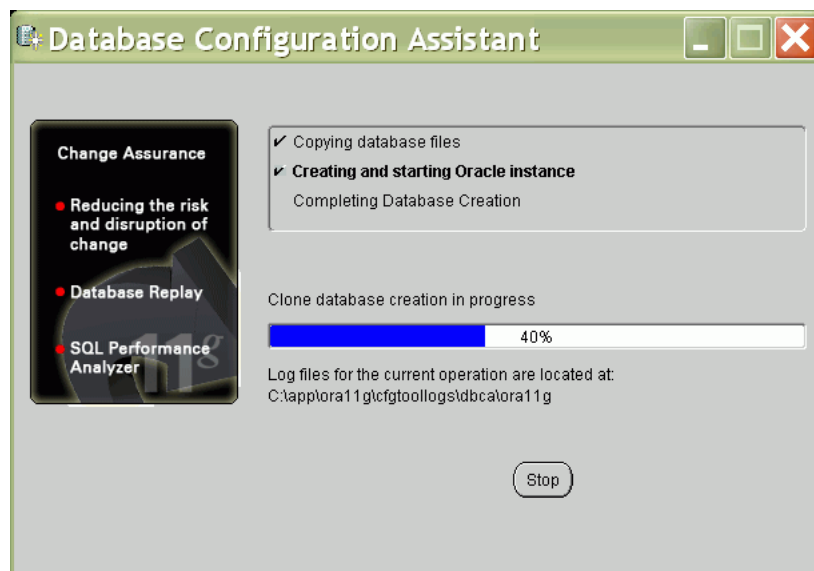


6. The Installer invokes the Configuration Assistants to configure and start the starter database.

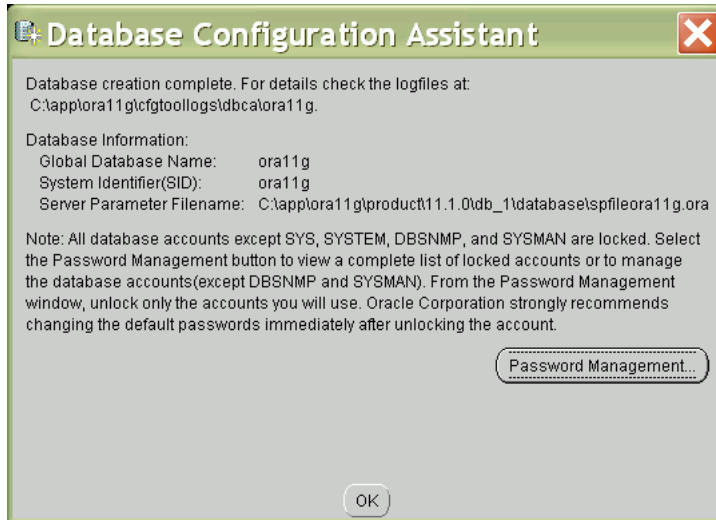


If the Configuration Assistants encounter an error, check the logs to determine the problem. You can choose to continue the installation and start the assistants manually later, or you can restart the installation. To continue the installation, click **Install**.

7. Database Configuration Assistant creates the starter database.



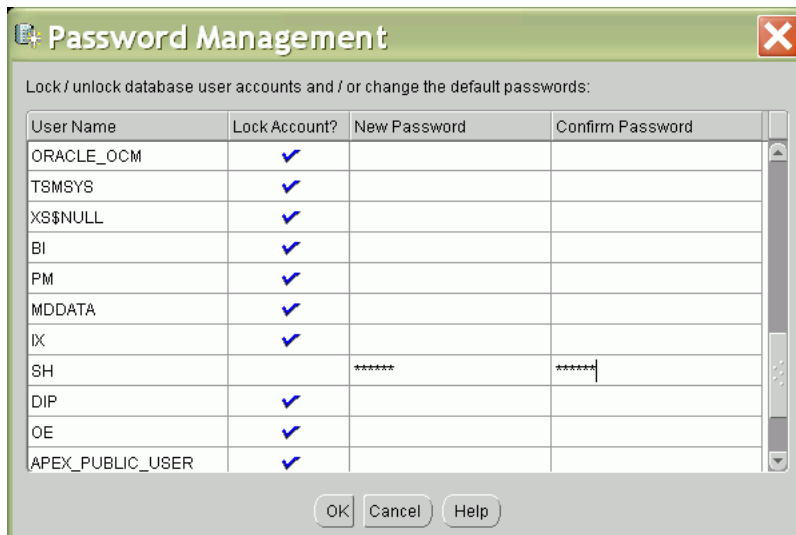
8. The Database Configuration Assistant page displays information about the starter database.



Click the **Password Management** button.

9. Unlock the *SYS*, *SYSTEM*, and *SH* accounts. Specify a password for *SH*. You can also change the passwords for *SYS* and *SYSTEM* if you wish. The password must have at least eight characters and include both alphabetic and numeric characters

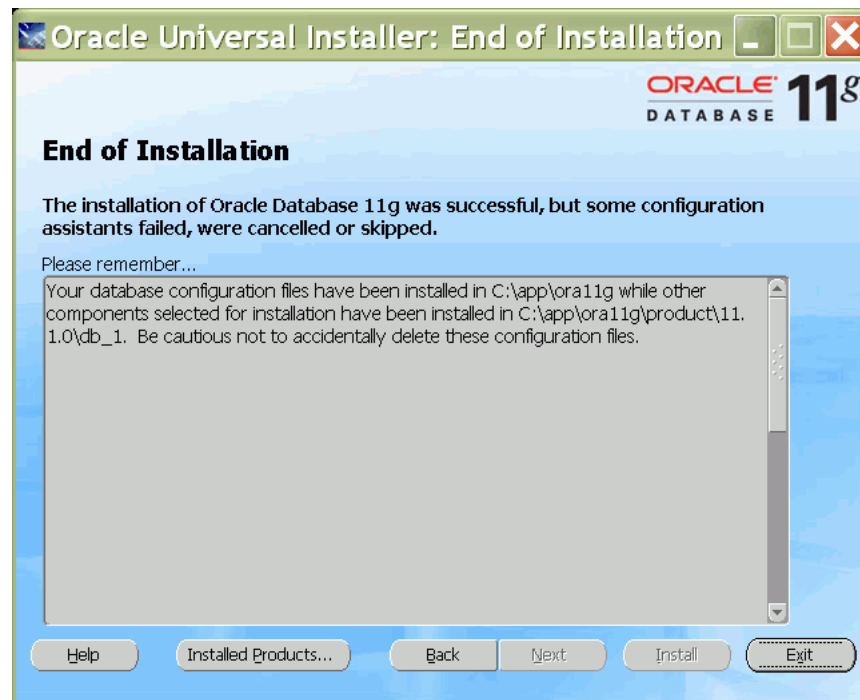
For details about specifying passwords, refer to *Oracle Database Security Guide*.



Click **OK** to return to the Database Configuration Assistant page.

On the Database Configuration Assistant page, click **OK**.

10. Click **EXIT** to exit the Installer.



Install Oracle Database Companion

The Oracle Data Mining sample programs are installed with Oracle Database Companion.

The Database Companion installation process copies the Oracle Data Mining sample programs, along with examples and demonstrations of other database features, to the \rdbms\demo subdirectory of the Oracle home directory.

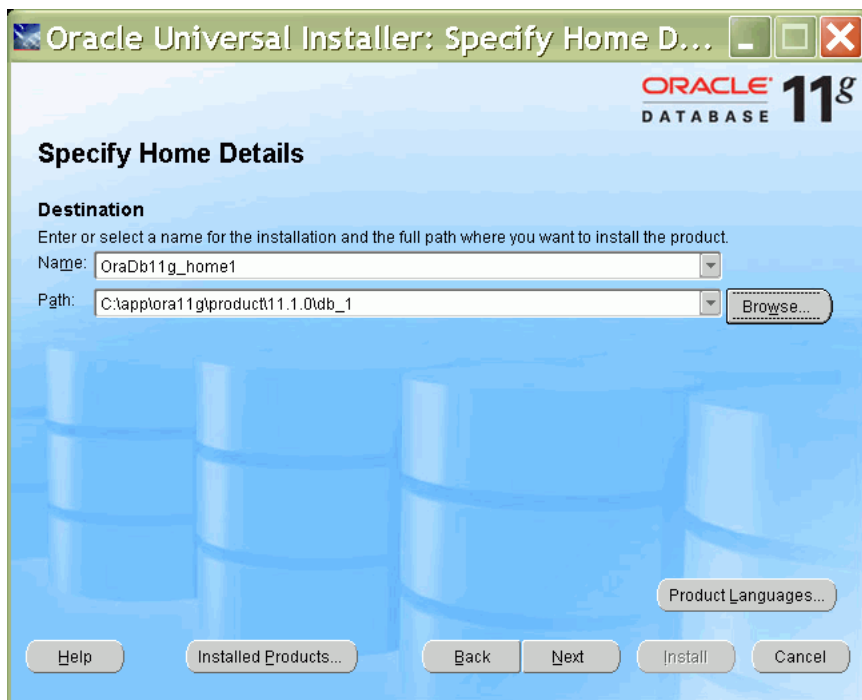
To install the Database Companion, perform these steps:

1. From the Companion installation directory, run `SETUP . EXE`.

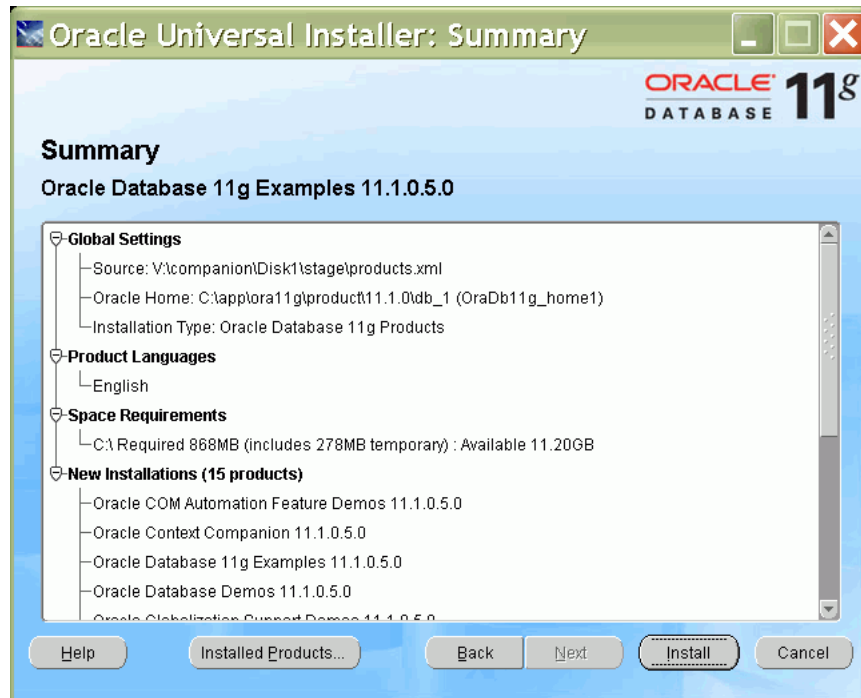
Oracle Universal Installer opens and displays the Welcome page. Click **Next** to advance to the next page.



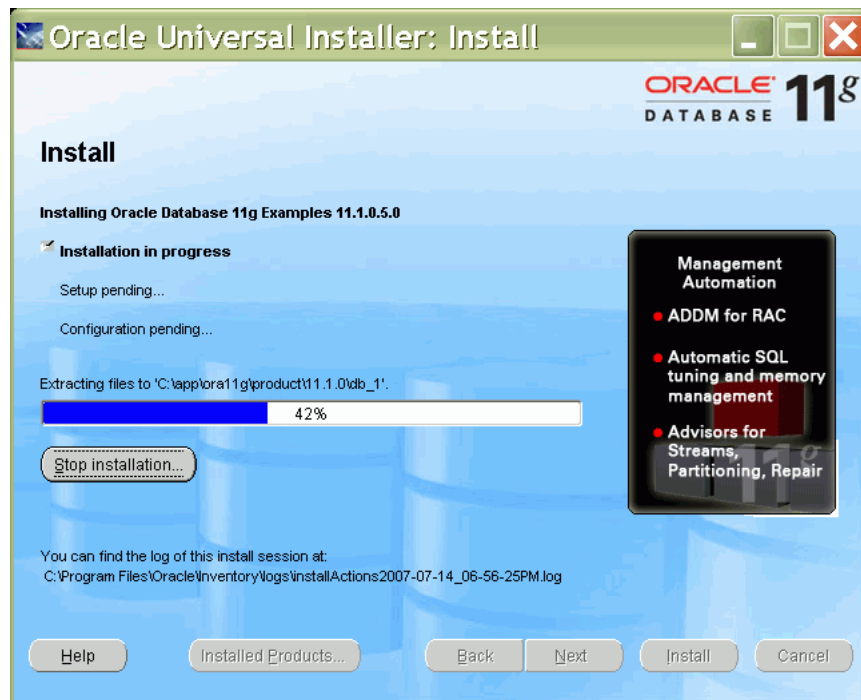
2. On the Specify Home Details page, specify the Oracle home directory in which you installed Oracle Database. Do not assume that the directory displayed by the Installer is correct.



3. On the Summary page, review the information and settings for your installation, then click **Install**.



- The Installer proceeds with the installation.



- On the End of Installation page, confirm that the installation was successful.



6. Click **Exit** to exit the Installer.

Create a Data Mining Demo User

To build and score Data Mining models, you must have an Oracle user ID with the appropriate privileges. Follow these instructions to create a demo user that has required privileges for running the sample programs and creating and scoring models within the user's schema.

See Also: [Chapter 4, "Users and Privileges for Data Mining"](#) to create data mining users that are capable of performing broader data mining tasks

Note: In the following sections, you will find references to the environment variable for the Oracle home directory on Windows (%ORACLE_HOME%). If the environment variable does not exist on your computer, you can create it. To add or change environment variables, open System in Control Panel and choose the **Advanced** tab.

1. Start SQL*Plus and login with system privileges.

You can launch SQL*Plus from the Windows Start menu. Choose the Oracle home menu item and the Application Development submenu.

```
Enter user-name: sys / as sysdba
Enter password: password
```

2. To create the user, type a command like the following.

```
CREATE USER dmuser IDENTIFIED BY password
DEFAULT TABLESPACE USERS
TEMPORARY TABLESPACE TEMP
```



```
QUOTA UNLIMITED ON USERS;
```

3. Run `dmshgrants.sql` to grant access to the SH schema. Several tables in SH are used by the Data Mining sample programs. Specify the Data Mining user name and the password to SH as parameters.

```
@ %ORACLE_HOME%\rdbms\demo\dmshgrants sh_password dmuser
```

4. Now connect to the database as the Data Mining user.

```
CONNECT dmuser
Enter password: password
```

5. Run `dmsh.sql` to populate the schema of the Data Mining user with tables, views, and other objects needed by the sample programs.

```
@ %ORACLE_HOME%\rdbms\demo\dmsh
COMMIT;
```

Once you have completed these steps, you can run the Data Mining sample programs whenever you log in to the database as the Data Mining demo user.

Run the Sample Programs

To locate the sample programs on your computer, navigate to the `rdbms\demo` subdirectory under Oracle home.

To display the Data Mining PL/SQL sample programs, search for the files that start with `dm` and end with `.sql`. (The list will include `dmsh.sql` and `dmshgrants.sql`, which are used to configure the Data Mining demo user ID.) The PL/SQL sample programs are listed in [Table 1–1](#).

Table 1–1 Sample PL/SQL Data Mining Programs

Program File	Algorithm	Mining Function or Task
<code>dmaidemo.sql</code>	Minimum Descriptor Length	Attribute Importance
<code>dmardemo.sql</code>	Apriori	Association
<code>dmdtdemo.sql</code>	Decision Tree	Classification
<code>dmdtxvlddemo.sql</code>	Decision Tree (cross validation)	Classification
<code>dmglcdem.sql</code>	Binary Logistic Regression (GLM)	Classification
<code>dmglrdem.sql</code>	Multivariate Linear Regression (GLM)	Regression
<code>dmkmdemo.sql</code>	<i>k</i> -Means	Clustering
<code>dmnbdemo.sql</code>	Naive Bayes	Classification
<code>dmnmdemo.sql</code>	Non-Negative Matrix Factorization	Feature Extraction
<code>dmocdemo.sql</code>	O-Cluster	Clustering
<code>dmsvcdem.sql</code>	Support Vector Machine	Classification
<code>dmsvodem.sql</code>	Support Vector Machine	Anomaly Detection
<code>dmsvrdem.sql</code>	Support Vector Machine	Regression
<code>dmtxtfe.sql</code>	Term extraction using Oracle Text	Text transformation for mining
<code>dmtxtnmf.sql</code>	Non-Negative Matrix Factorization	Text mining using NMF
<code>dmtxtsvm.sql</code>	Support Vector Machine	Text mining using SVM

In the same directory, search for the files that start with `dm` and end with `.java` to display the Java samples. The Java sample programs are listed in [Table 1–2](#).

Table 1–2 Sample Java Data Mining Programs

Program File	Algorithm	Mining Function or Task
<code>dmaidemo.java</code>	Minimum Description Length	Attribute importance
<code>dmapplydemo.java</code>	Naive Bayes	Illustrate scoring methods
<code>dmardemo.java</code>	Apriori	Association
<code>dmexpimpdemo.java</code>	NA	Model Export/Import
<code>dmglcdemo.java</code>	Binary Logistic Regression (GLM)	Classification
<code>dmglrdemo.java</code>	Multivariate Linear Regression (GLM)	Regression
<code>dmkmdemo.java</code>	<i>k</i> -Means	Clustering
<code>dmnbdemo.java</code>	Naive Bayes	Classification
<code>dmnmdemo.java</code>	Non-Negative Matrix Factorization	Feature extraction
<code>dmocdemo.java</code>	O-Cluster	Clustering
<code>dmpademo.java</code>	Automated predict and explain	Predictive Analytics
<code>dmsvcdemo.java</code>	Support Vector Machine	Classification
<code>dmsvodemo.java</code>	Support Vector Machine (one class)	Classification
<code>dmsvrdemo.java</code>	Support Vector Machine	Regression
<code>dmtreedemo.java</code>	Decision Tree	Classification
<code>dmtxtnmfdemo.java</code>	Non-Negative Matrix Factorization	Text mining with NMF
<code>dmtxtsvmdemo.java</code>	Support Vector Machine	Text mining with SVM classification
<code>dmxfdemo.java</code>	Binning, clipping, and normalization	Data Transformations

View the Source Code

You will learn a great deal about the Data Mining APIs by investigating the source code of the sample programs. The programs illustrate typical approaches to data preparation, algorithm selection, algorithm tuning, testing, and scoring. All the programs include extensive comments to help you understand what the code is doing.

You can view the source code simply by opening the files in a text editor.

Run the PL/SQL Sample Programs

Now that you have a user ID with the required privileges and a schema populated with the required objects, you can run the sample programs. Each program creates a Data Mining model.

While the program is running, it displays the code and the program output.

You can run the sample programs as many times as you wish. The programs clean up the results of the previous run before executing the current run.

To run the PL/SQL programs:

1. Start SQL*Plus and login as the Data Mining user.

```
Enter user-name: dmuser
Enter password: password
```

2. Run the program by specifying "@" followed by the fully-qualified path of the program. This example executes the program `dmnbdemo.sql`, which creates a Naive Bayes model.

```
SQL>@ %ORACLE_HOME%\rdbms\demo\dmnbdemo
```

Prepare to Run the Java Programs

Before you can run the Java programs, you must set up your Java environment and compile the programs. You can do this in an Integrated Development Environment such as Oracle JDeveloper, or you can execute the following commands at the operating system prompt.

1. Check that the version of Java you are using is 1.5 or higher. You can execute the following in a command window to check the version of Java.

```
>java -version
```

2. Add `%ORACLE_HOME%\jdk\bin\` to your PATH variable before the paths of any other Java versions.
3. Add the following Data Mining JAR files to your Windows CLASSPATH:

```
%ORACLE_HOME%\rdbms\jlib\jdm.jar
%ORACLE_HOME%\rdbms\jlib\ojdm_api.jar
%ORACLE_HOME%\rdbms\jlib\xdb.jar
%ORACLE_HOME%\jdbc\lib\ojdbc5.jar
%ORACLE_HOME%\oc4j\j2ee\home\lib\connector.jar
%ORACLE_HOME%\jlib\orai18n.jar
%ORACLE_HOME%\jlib\orai18n-mapping.jar
%ORACLE_HOME%\lib\xmlparserv2.jar
```

4. Compile the programs listed in [Table 1-2](#). To use the JAVAC executable, open a command window and go to `\rdbms\demo` in Oracle home.

```
>javac program_name.java
```

For example:

```
>javac dmnbdemo.java
```

If JAVAC is not found, then check the value of the PATH variable.

Run the Java Programs

You can run a Java program from the operating system prompt with a command like this:

```
>java program_name host_name:port_number:database_identifier user password
```

View the Models Created by the Sample Programs

You can query the `USER_MINING_MODELS` view to list the models in your schema.

```
SQL> set linesize 100
```

```
SQL> SELECT model_name, mining_function, algorithm from user_mining_models;
```

MODEL_NAME	MINING_FUNCTION	ALGORITHM
AI_SH_SAMPLE	ATTRIBUTE_IMPORTANCE	MINIMUM_DESCRIPTION_LENGTH

AR_SH_SAMPLE

ASSOCIATION_RULES

APRIORI_ASSOCIATION_RULES

This example shows that there are two mining models in your schema. The model name, mining function, and algorithm are displayed. To find all the columns defined in a view, use a DESCRIBE command.

```
SQL> DESCRIBE user_mining_models
```

You can query the USER_MINING_MODEL_ATTRIBUTES and USER_MINING_MODEL_SETTINGS views to obtain information about the attributes and settings for the models in your schema.

Connecting to a Database for Data Mining

This chapter explains, in general terms, how to connect to a remote Oracle database to perform data mining and other database activities.

Note: Oracle Database Online Documentation Library contains simple step-wise instructions to help you connect to and interact with an Oracle database. Here are two good places to start:

- *Oracle Database 2 Day DBA*
 - *Oracle Database 2 Day Developer's Guide*
-
-

This chapter contains the following sections:

- [Your User ID](#)
- [Connect to a Local Database](#)
- [Connect to a Remote Database](#)
- [Create a Net Service Name](#)

Your User ID

To connect to Oracle Database, you must log on with a user ID and password. To perform data mining activities in the database, you must log on with a user ID that has been granted the necessary database privileges.

If you simply wish to run the Data Mining sample programs, see "[Create a Data Mining Demo User](#)" on page 1-10.

If you wish to perform broader data mining activities, refer to [Chapter 4, "Users and Privileges for Data Mining"](#).

Connect to a Local Database

In [Chapter 1](#), you learned how to install Oracle Database on your personal computer. When you install Oracle Database locally, you do not need to install separate client software to connect to the database.

SQL*Plus and other client tools are installed with Oracle Database. The SQL*Plus executable is installed in the `bin` directory under Oracle home. You can start SQL*Plus from the system command prompt or from the Windows Start menu.

You can use SQL*Plus to connect to your local database by simply specifying the global database name and providing a user name and password.

You can also use SQL*Plus to connect to a remote database by specifying a user name, password, and a Net Service Name. See ["Create a Net Service Name"](#) on page 2-3.

Connect to a Remote Database

If you do not have a local database, you must install Oracle Client to obtain SQL*Plus and other tools for connecting to an Oracle database.

Install Oracle Client

Use the following steps to install Oracle Client on a Windows platform:

1. From the Client installation directory, run `SETUP.EXE`.
Oracle Universal Installer opens and displays the Welcome page. Click **Next**.
2. On the Select Installation Type page, choose **Administrator** and click **Next**.
3. On the Specify Home Details page, provide the path of a home directory for the Oracle Client installation.
4. On the Product-Specific Prerequisite page, verify that all checks succeeded. If any checks failed, then you must correct the problem before proceeding.
5. On the Summary page, click the **Install** button.
6. The Installer displays the progress of the installation. When you click **Next**, the Configuration Assistants page is displayed.
7. Oracle Net Configuration Assistant starts and displays the Welcome page. Choose **Perform Typical Configuration**.

Oracle Net Configuration Assistant creates a simple connection using the Easy Connect naming method. This method enables clients to connect to a remote database server without any configuration. Clients specify a SQL `CONNECT` statement using a simple TCP/IP address, identified by a host name and an optional port number and service name.

```
CONNECT username@"host[:port] [/servicename]"
```

Note: The system will prompt for a password.

8. When the Oracle Net Configuration process is complete, click **Finish**.

See Also: *Oracle Database Client Installation Guide for Microsoft Windows*

Perform Data Mining on a Remote Computer

You can connect to a remote instance of Oracle Database and perform data mining activities as long as:

- SQL*Plus (or another client tool such as SQL Developer) has been installed on your computer.
- The system administrator has created a user ID for you in the remote database.
- The system administrator has granted the privileges to your user ID to allow the data mining activities that you will perform.

- The system administrator has provided you with the Oracle Net Service Name for connecting to the remote database.

Run the Data Mining Sample Programs on a Remote Computer

If you wish to run the Data Mining sample programs on a remote computer:

1. Go to the Sample Code page of Oracle Technology Network:
http://www.oracle.com/technology/sample_code/index.html
2. Scroll down to **Sample Code — Data Warehousing & Business Intelligence** and choose **Oracle Data Mining**.
3. Unzip the sample programs to a directory on your computer.
4. Ask the system administrator to install the SH schema in the remote database, if it has not already been installed.
5. Ask the system administrator to run `dmshgrants.sql` to grant the appropriate privileges to your user ID.
6. Start SQL*Plus and log in to the remote database as the Data Mining user. Run `dmsh.sql` to populate your schema with objects used by the sample programs. (See "Create a Data Mining Demo User" on page 1-10 for information about `dmshgrants.sql` and `dmsh.sql`.)
7. Follow the instructions in "Run the Sample Programs" on page 1-11.

Create a Net Service Name

A Net Service Name specifies the name of a database (the global database name), the name of the computer (host) on which it is installed, and the port number where the database listens for requests from a client.

There are several ways to create a Net Service Name. One method is the Easy Connect naming method that can be performed during the installation of Oracle Client (described in step 7 of "Connect to a Remote Database" on page 2-2). You can also use Oracle Net Configuration Assistant to create a Net Service Name..

See Also: "Configuring the Network Environment" in *Oracle Database 2 Day DBA* to learn about creating and specifying a Net Service Name

Upgrading Oracle Data Mining

This chapter explains how to upgrade or downgrade your database.

See Also: ["What's New in Oracle Data Mining Administration"](#) on page -xi for a summary of changes in Oracle Data Mining 11g that affect database administration

This chapter contains the following sections:

- [Upgrading Oracle Data Mining](#)
- [Downgrading Oracle Data Mining](#)

Upgrading Oracle Data Mining

Oracle Data Mining upgrade is fully integrated with the upgrade of Oracle Database. In Oracle 11g, Data Mining metadata is migrated from `DMSYS` to `SYS` during the upgrade of Oracle Database.

Data Mining models are also upgraded during the upgrade of Oracle Database. The upgraded models are available for use in the new upgraded environment. Upgraded models will continue to work as they did in the prior release. All new mining functionality in 11g can be used in the upgraded environment.

In Oracle 11g, the `CREATE MINING MODEL` privilege must be granted to Data Mining user accounts. This privilege is required for creating Data Mining models in 11g. Refer to [Chapter 4, "Users and Privileges for Data Mining"](#) for more information.

See Also: *Oracle Database Upgrade Guide* for information about upgrading Oracle Database

Supported Version Upgrades

Oracle Data Mining 10.2, including both metadata and models, can be upgraded to 11g.

Models created by the Oracle Data Mining PL/SQL API can be upgraded from 10.1 to 11g. Models created by the Oracle Data Mining Java API cannot be upgraded from 10.1 to 11g. The 10.1 version of the Java API was no longer supported in Oracle Data Mining 10.2.

Model upgrade from 9.2 to 11g is not supported.

Using Oracle Database Upgrade Assistant

You can use Oracle Database Upgrade Assistant to upgrade a database to 11g. During the upgrade, all Data Mining metadata that previously existed in the `DMSYS` schema is created in `SYS`. After the upgrade, all Data Mining metadata will reside in `SYS` and will no longer reside in `DMSYS`. Data Mining models, which reside in the owner's schema, will also be upgraded to 11g.

On Windows platforms, you can use the Database Upgrade Assistant graphical tool to upgrade to 11g. To start the Upgrade Assistant:

1. Go to the Windows **Start** menu and choose the Oracle home directory.
2. Choose the **Configuration and Migration Tools** menu.
3. Launch the **Upgrade Assistant**.

On Linux platforms, run the `DBUA` utility to upgrade Oracle Database.

After upgrading, check the upgrade log file and `DBA_REGISTRY` to ensure that the upgrade process completed successfully. Also check the `DBA_MINING_MODELS` view in the upgraded database. The newly upgraded mining models should be listed in this view.

After you have verified the upgrade and confirmed that there will be no need to downgrade, you should set the initialization parameter `COMPATIBLE` to `11.1`. At that point, you can drop the `DMSYS` schema from the database. Once `DMSYS` is removed, the `DBA_REGISTRY` will no longer list Oracle Data Mining as a component.

Note: After upgrading to 11g, you can no longer switch to the Data Mining Scoring Engine (DMSE). The Scoring Engine does not exist in 11g.

Exporting and Importing from a Dump File

If you wish, you can use a less automated approach to upgrading Oracle Data Mining. You can export the models created in a previous version of Oracle Database and import them into a new 11g database.

To export your Data Mining models to a dump file, follow the instructions in [Chapter 6](#) on import/export.

Before importing from the dump file, run the `DMEIDMSYS` script to create the `DMSYS` schema in the 11g database.

```
SQL>CONNECT / as sysdba;
SQL>@ORACLE_HOME/rdbms/admin/dmeidmsys.sql
SQL>EXIT;
```

Note: The `TEMP` tablespace must already exist in the 11g database. The `DMEIDMSYS` script uses `TEMP` and `SYSAUX` to create the `DMSYS` schema.

To import the dump file into the database:

```
%ORACLE_HOME/bin/impdp system/password file_name = .....
SQL>CONNECT / as sysdba;
SQL>EXECUTE dmp_sys.upgrade_models('11.0.0');
SQL>ALTER SYSTEM FLUSH SHARED_POOL;
SQL>EXIT;
```

If you shutdown the database before operating on the upgraded mining models, this will also flush the shared pool.

After the import is complete, you can drop the DMSYS schema from the database. Once DMSYS is removed, the DBA_REGISTRY will no longer list Oracle Data Mining as a component.

Downgrading Oracle Data Mining

Before downgrading the database back to the previous version, ensure that there are no 11g mining models in the upgraded database. Issue the following SQL statement in SYS to verify this:

```
SQL>SELECT o.name FROM sys.model$ m, sys.obj$ o
        WHERE m.obj#=o.obj# AND m.version=2;
```

If there are any 11g mining models in the database, you must manually delete them using the DBMS_DATA_MINING.DROP_MODEL routine before downgrading the database. If you do not do this, the database downgrade process will be aborted.

Users and Privileges for Data Mining

This chapter explains how to create data mining users and control their mining activities through the use of database privileges.

See Also: *Oracle Database 2 Day + Security Guide* for information about Oracle Database security

This chapter contains the following topics:

- [Creating Data Mining Users](#)
- [System Privileges for Mining Models](#)
- [Object Privileges on Mining Models](#)

Creating Data Mining Users

Data Mining should be performed by a user that is specifically configured for data mining activities.

The following SQL statement creates a database user and password. The user will use a tablespace called `default_tablespace` by default; its temporary tablespace is called `temp_tablespace`. Tablespaces for Data Mining users should be assigned according to standard DBA practices, depending on system load and system resources.

```
CREATE USER dmuser IDENTIFIED BY password
  DEFAULT TABLESPACE default_tablespace
  TEMPORARY TABLESPACE temp_tablespace
  QUOTA UNLIMITED on default_tablespace;
```

You must be logged on as a system administrator to create users and grant privileges.

Note: To create a user that can use the Data Mining sample programs, you need to run two configuration scripts as described in "[Create a Data Mining Demo User](#)" on page 1-10.

Privileges Required for Data Mining

Oracle Corporation recommends a set of basic privileges for data mining. The following GRANT statements grant these privileges to a user named `dmuser`.

```
GRANT create mining model TO dmuser;
GRANT create procedure TO dmuser;
GRANT create session TO dmuser;
```

```
GRANT create table TO dmuser;
GRANT create sequence TO dmuser;
GRANT create view TO dmuser;
GRANT create job TO dmuser;
GRANT create type TO dmuser;
GRANT create synonym TO dmuser;
```

The `CREATE MINING MODEL` privilege is required for creating models. The other privileges may not be required for some mining activities, however it is prudent to grant them all as a group.

`SELECT` access to the data being mined is always required. The following example allows `dmuser` to mine data in the `CUSTOMERS` table in the `SH` schema.

```
GRANT SELECT ON customers TO dmuser;
```

`EXECUTE` access to the Oracle Text package `ctxsys.ctx_ddl` is required for text mining.

```
GRANT EXECUTE ON ctxsys.ctx_ddl TO dmuser;
```

Additional access rights are required for exporting and importing mining models, as described in ["Exporting and Importing Mining Models"](#) on page 6-5.

Additional system and object privileges described in the following sections can be used to enable or restrict particular mining activities.

System Privileges for Mining Models

A **system privilege** confers the right to perform a particular action in the database or to perform an action on a type of schema objects. For example, the privileges to create tablespaces and to delete the rows of any table in a database are system privileges.

To grant a system privilege, you must either have been granted the system privilege with the `ADMIN OPTION` or have been granted the `GRANT ANY PRIVILEGE` system privilege.

The system privileges listed in [Table 4-1](#) are required for performing specific operations on mining models.

Table 4-1 System Privileges for Mining Model

System Privilege	Allows you to
<code>CREATE MINING MODEL</code>	Create mining models in your own schema.
<code>CREATE ANY MINING MODEL</code>	Create mining models in any schema.
<code>ALTER ANY MINING MODEL</code>	Change the name or cost matrix of any mining model in any schema.
<code>DROP ANY MINING MODEL</code>	Drop any mining model in any schema.
<code>SELECT ANY MINING MODEL</code>	Apply a mining model in any schema, also view model details in any schema.
<code>COMMENT ANY MINING MODEL</code>	Add a comment to any mining model in any schema.
<code>AUDIT ANY</code>	Generate an audit trail for any mining model (or any object) in any schema.

Object Privileges on Mining Models

An object privilege confers the right to perform a particular action on a specific schema object. For example, the privilege to delete rows from the SH . PRODUCTS table is an example of an object privilege.

A user automatically has all object privileges for schema objects contained in his or her schema. A user can grant any object privilege on any schema object he or she owns to any other user or role.

Table 4–2 *Object Privileges for Data Mining*

Object Privilege	Allows you to
ALTER MINING MODEL	Change the name or cost matrix of the specified mining model object.
SELECT MINING MODEL	Apply or view the specified mining model object.

Installing the Spreadsheet Add-In for Predictive Analytics

This chapter provides instructions for installing Oracle Spreadsheet Add-In for Predictive Analytics.

This chapter contains the following topics:

- [About the Spreadsheet Add-In](#)
- [Requirements](#)
- [Install the Spreadsheet Add-In](#)

About the Spreadsheet Add-In

The Oracle Spreadsheet Add-In for Predictive Analytics adds predictive analytics features to Microsoft Excel. Using simple "one click" data mining for `PREDICT`, `EXPLAIN`, and `PROFILE` operations, Excel users can mine data stored in Excel or in Oracle Database.

Predictive analytics provides automated methodologies that simplify data mining.

See Also: *Oracle Data Mining Concepts* for information about Oracle predictive analytics

Requirements

The Spreadsheet Add-In requires access to an instance of Oracle Database installed with the Data Mining option. The database can be installed locally on your computer or it can be installed on a remote host.

Your user ID must have the database and object permissions described in [Chapter 4](#). If you want to use the sample data provided with the sample programs, the `SH` schema must be installed in the database and your user ID must be enabled as described in "[Create a Data Mining Demo User](#)" on page 1-10.

Oracle Spreadsheet Add-In for Predictive Analytics requires the presence of Oracle Objects for OLE and Oracle Net Configuration Assistant. To obtain these components, you must install Oracle Client on the computer where Excel and the Spreadsheet Add-In are installed.

In summary, Oracle Spreadsheet Add-In requires:

- Microsoft Excel 2000 or later
- Oracle Objects for OLE, which must be installed with Oracle Client

- An Oracle Net Service Name

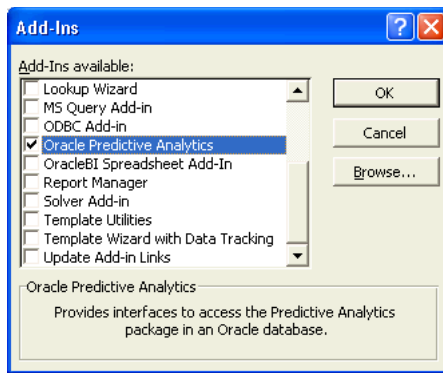
Install the Spreadsheet Add-In

Follow these steps to install the Spreadsheet Add-In:

1. Download the Spreadsheet Add-In from the Oracle Data Mining Web site at <http://www.oracle.com/technology/products/bi/odm/index.html>.
2. Open the ZIP file and extract the file named `Predictive_Analytics.xla` to the Microsoft Office Library directory. The library has a path such as this one:

`C:\Program Files\Microsoft Office\Office\Library`

3. Open Excel and click **Tools > Add-Ins**.
4. Select **Oracle Predictive Analytics** from the Add-Ins dialog box, as shown in the following figure.



The OraclePA menu is added to the Excel toolbar.

5. Install Oracle Client as described in "[Install Oracle Client](#)" on page 2-2.
6. Create an Oracle Net Service Name to manage the connection between Excel and Oracle Database. You can use the Easy Connect Naming method described in the installation instructions for Oracle Client, or you can use Oracle Net Configuration Assistant.

The Spreadsheet Add-In uses features of Oracle Data Mining. Whether the data to be mined is stored in Excel or in Oracle Database, an Oracle Net Service Name is required for accessing the data mining functionality.

Even if the Oracle Database is local, you still need to create a Net Service Name for the Spreadsheet Add-In.

See Also:

- See "[Create a Net Service Name](#)" on page 2-3
- "Configuring the Network Environment" in *Oracle Database 2 Day DBA*

7. In Excel, click **OraclePA > Connect**.
8. On the Connect dialog box, select a service name from the drop-down list, and type in your database user name and password.

Mining Model Schema Objects

In this chapter, you will learn how to find information about mining models in the data dictionary and how to perform various operations on mining models.

See Also: [Chapter 4, "Users and Privileges for Data Mining"](#) for information about system and object privileges associated with mining model objects.

This chapter contains the following topics:

- [Obtaining Information from the Data Dictionary](#)
- [Adding a Comment to a Mining Model](#)
- [Auditing Mining Models](#)
- [Exporting and Importing Mining Models](#)

Obtaining Information from the Data Dictionary

Mining models are database schema objects. They can be queried in the ALL, DBA, and USER data dictionary views.

The data dictionary views in [Table 6–1](#) reveal information about mining models created by Oracle Data Mining.

Table 6–1 Oracle Data Mining Data Dictionary Views

ALL_ Views	DBA_ Views	USER_ Views
ALL_MINING_MODELS	DBA_MINING_MODELS	USER_MINING_MODELS
ALL_MINING_MODEL_ATTRIBUTES	DBA_MINING_MODEL_ATTRIBUTES	USER_MINING_MODEL_ATTRIBUTES
ALL_MINING_MODEL_SETTINGS	DBA_MINING_MODEL_SETTINGS	USER_MINING_MODEL_SETTINGS

Obtaining Information about Mining Models

You can query the ALL_MINING_MODELS data dictionary view to obtain information about all accessible mining model objects. USER and DBA versions of this view are also available.

```
SQL> DESCRIBE all_mining_models
```

```
Name                                     Null?    Type
-----
OWNER                                     NOT NULL VARCHAR2 (30)
MODEL_NAME                               NOT NULL VARCHAR2 (30)
MINING_FUNCTION                           VARCHAR2 (30)
```

ALGORITHM		VARCHAR2 (30)
CREATION_DATE	NOT NULL	DATE
BUILD_DURATION		NUMBER
MODEL_SIZE		NUMBER
COMMENTS		VARCHAR2 (4000)

The COMMENTS column contains comments created by SQL COMMENT, if they exist. See ["Adding a Comment to a Mining Model"](#) on page 6-3.

See Also:

Oracle Data Mining Concepts for information about mining functions and algorithms

Oracle Data Mining Application Developer's Guide for more information about ALL_MINING_MODELS

Obtaining Information about Mining Model Attributes

You can query the ALL_MINING_MODEL_ATTRIBUTES data dictionary view to obtain information about all accessible mining model attributes. USER and DBA versions of this view are also available.

```
SQL> DESCRIBE all_mining_model_attributes
```

Name	Null?	Type
OWNER	NOT NULL	VARCHAR2 (30)
MODEL_NAME	NOT NULL	VARCHAR2 (30)
ATTRIBUTE_NAME	NOT NULL	VARCHAR2 (30)
ATTRIBUTE_TYPE		VARCHAR2 (11)
DATA_TYPE		VARCHAR2 (12)
DATA_LENGTH		NUMBER
DATA_PRECISION		NUMBER
DATA_SCALE		NUMBER
USAGE_TYPE		VARCHAR2 (8)
TARGET		VARCHAR2 (3)

Note: The attributes listed in this data dictionary view are the physical columns in the build data that were used to construct the model. Some or all of these columns should be present for scoring. These **data attributes** are referred to as the **model signature**.

The term *attribute* is more accurately used to designate the numericals and categoricals derived from the data attributes for manipulation by the algorithm. These **model attributes** may or may not correspond to data attributes, depending on transformations and on whether or not the column is nested. The model attributes can be viewed in the model details (GET_MODEL_DETAILS functions).

See Also:

Oracle Data Mining Application Developer's Guide for more information about attributes

Oracle Data Mining Application Developer's Guide for more information about ALL_MINING_MODEL_ATTRIBUTES

Obtaining Information about Mining Model Settings

You can query the ALL_MINING_MODEL_SETTINGS data dictionary view to obtain information about all accessible mining model settings. USER and DBA versions of this view are also available.

```
SQL> DESCRIBE all_mining_model_settings
```

Name	Null?	Type
OWNER	NOT NULL	VARCHAR2 (30)
MODEL_NAME	NOT NULL	VARCHAR2 (30)
SETTING_NAME	NOT NULL	VARCHAR2 (30)
SETTING_VALUE		VARCHAR2 (4000)
SETTING_TYPE		VARCHAR2 (7)

Model settings can be specified in a settings table used in the model build. The settings all have default values. The defaults are used when they are not overridden by settings specified in the settings table, or when there is no settings table.

See Also:

Oracle Database PL/SQL Packages and Types Reference for descriptions of model settings

Oracle Data Mining Concepts for details about algorithm settings

Oracle Data Mining Application Developer's Guide for more information about ALL_MINING_MODEL_SETTINGS

Adding a Comment to a Mining Model

You can associate a comment with a mining model using a SQL COMMENT statement.

```
COMMENT ON MINING MODEL schema_name.model_name IS string;
```

Note: To add a comment to a model in another schema, you must have the COMMENT ANY MODEL system privilege.

To drop a comment, set it to the empty ' ' string.

The following statement adds a comment to the model DT_SH_CLAS_SAMPLE in your own schema.

```
SQL> COMMENT ON mining model dt_sh_clas_sample IS
      'Decision Tree model predicts promotion response';
```

You can view the comment by querying the catalog view USER_MINING_MODELS.

```
SQL> COLUMN comments FORMAT a22
```

```
SQL> SELECT model_name, mining_function, algorithm, comments FROM user_mining_models;
```

MODEL_NAME	MINING_FUNCTION	ALGORITHM	COMMENTS
DT_SH_CLAS_SAMPLE	CLASSIFICATION	DECISION_TREE	Decision Tree model predicts promotion response

To drop this comment from the database, issue the following statement:

```
SQL> COMMENT ON mining model dt_sh_clas_sample '';
```

Auditing Mining Models

You can use the SQL auditing system to track operations on data mining models.

Note: To audit a mining model in another schema, you must have the `AUDIT ANY` system privilege.

Enabling Auditing in the Database

The database initialization parameter `AUDIT_TRAIL` controls auditing capabilities in the database. To enable auditing, set `AUDIT_TRAIL` to `DB`, `DB_EXTENDED`, or `OS`.

Set `AUDIT_TRAIL` to `NONE` to prevent auditing information from being recorded. By default, `AUDIT_TRAIL` is set to `NONE`.

Opening an Audit Trail on Mining Models

Use the SQL `AUDIT` statement to open an auditing trail on a data mining model.

```
AUDIT {operation|ALL} ON mining model schema_name.model_name;
```

You can track the following operations on mining models.

Audit Operation	Description
AUDIT	Generate an audit trail for a mining model
COMMENT	Add a comment to a mining model
GRANT	Give permission to a user to access the model
RENAME	Change the name of the model
SELECT	Apply the model or view its signature.

For example, this statement generates an audit trail for all `GRANT` operations on the model `NB_SH_CLAS_SAMPLE` in the `DMUSER` schema.

```
SQL> AUDIT GRANT ON mining model dmuser.nb_sh_clas_sample;
```

This statement generates an audit trail for all operations on the same model.

```
SQL> AUDIT GRANT,AUDIT,COMMENT,RENAME,SELECT
      ON mining model dmuser.nb_sh_clas_sample;
```

You can refine the criteria for auditing with the following additional semantics.

```
AUDIT {operation|ALL} ON MINING MODEL schema_name.model_name
      [BY [SESSION|ACCESS]]
      [WHENEVER [NOT] SUCCESSFUL];
```

Specify `BY SESSION` if you want Oracle Database to write a single record for all operations of the same type on each mining model in the same session. Specify `BY ACCESS` if you want Oracle Database to write one record for each audited operation.

Closing the Audit Trail

Use the `NOAUDIT` statement to stop one or more auditing operations previously enabled by the `AUDIT` statement.

```
NOAUDIT {operation| ALL} ON MINING MODEL model_name
```

[WHENEVER [NOT] SUCCESSFUL]] ;

Viewing the Audit Trail

For each audited operation, Oracle Database produces an audit record containing:

- The name of the user performing the operation
- The type of operation
- The object involved in the operation
- The date and time of the operation

Several data dictionary views present auditing information. Some examples are:

- `DBA_AUDIT_OBJECT` displays audit trail records for all objects in the database.
- `USER_AUDIT_OBJECT` displays audit trail records for all objects accessible to the current user
- `DBA_OBJ_AUDIT_OPTS` describes auditing options for all objects in the database.
- `USER_OBJ_AUDIT_OPTS` describes auditing options for all objects owned by the current user.

Note: The Oracle Database auditing system is a powerful, highly configurable tool for tracking operations on schema objects. Refer to the following manuals for more information:

- *Oracle Database SQL Language Reference* for documentation of the `AUDIT` and `NOAUDIT` statements
 - *Oracle Database Reference* for documentation of the `AUDIT_TRAIL` initialization parameter and the data dictionary views for querying the database audit trail.
 - *Oracle Database Security Guide* for a comprehensive discussion of database auditing.
-
-

Exporting and Importing Mining Models

You can export data mining models to flat files to back up work in progress or to move models to a different instance of Oracle Database Enterprise Edition (such as from a development database to a production database). All methods for exporting and importing models are based in Oracle Data Pump technology.

Oracle Data Pump consists of two command-line clients and two PL/SQL APIs. The command-line clients, `EXPDP` and `IMPDP`, provide an easy-to-use interface to the Data Pump export and import utilities. The Data Mining APIs also use the Data Pump export and import utilities.

You can export and import models at different levels, depending on your access rights in the database:

- **Database.** When a DBA exports a full database using `EXPDP`, all data mining models in the database are exported. The `IMPDP` utility imports all the models with the other objects in the database.
- **Schema.** When a DBA or an individual user exports a schema using `EXPDP`, all the data mining models in the schema are exported. Likewise, `IMPDP` imports all the models with the other objects in the schema.

- **Models Only.** The Data Mining APIs contain utilities for exporting and importing mining models. You can select all the mining models in a schema or mining models that match specific criteria.

You cannot select individual models using `EXPDP` or `IMPDP`.

The Data Pump export utility writes the tables and metadata that constitute a model to a dump file set, which consists of one or more files. The Data Pump import utility retrieves the tables and metadata from the dump file and restores them to the target database.

See Also:

- *Oracle Database Utilities* for a complete discussion of Oracle Data Pump and the `expdp` and `impdp` utilities
- *Oracle Database PL/SQL Packages and Types Reference* for detailed information about the export and import procedures in the `DBMS_DATA_MINING` package
- *Oracle Data Mining Java API Reference* for information about the export and import classes in the Oracle Data Mining Java API

Prerequisites

To export and import Data Mining models, you must have read and write access to a directory object, and you may need additional database permissions.

Directory Objects

A directory object is a logical name in the database for a physical directory on the host computer. Without read and write access to a directory object, you cannot access the host computer file system from within Oracle Database.

You must have the `CREATE ANY DIRECTORY` privilege to create directory objects.

The following SQL command creates, or re-creates if it already exists, a directory object named `DMTEST`. The file system directory (in this example, `C:\ORACLE\PRODUCT\11.1.0\DMINING`) must already exist and have shared read/write access rights granted by the operating system.

```
CREATE OR REPLACE DIRECTORY dmtest AS 'c:\oracle\product\11.1.0\dmning';
```

This SQL command gives user `DMUSER` both read and write access to `DMTEST`.

```
GRANT ALL ON DIRECTORY dmtest TO dmuser;
```

For more information about creating database directories, refer to the `CREATE DIRECTORY` and `GRANT` commands in the *Oracle Database SQL Language Reference*.

Additional Database Privileges

You may need special privileges in the database to take full advantage of all Data Pump features, such as importing models and other objects into a different schema. These privileges are granted by the `EXP_FULL_DATABASE` and `IMP_FULL_DATABASE` roles, which are only available to privileged users (such as `SYS` or a user with the `DBA` role).

You do not need these roles to export models from your own schema. To import models, you must have the same database privileges as the user who created the dump file set. Otherwise, a `DBA` with full system privileges must import the models.

PL/SQL APIs for Exporting and Importing Models

The DBMS_DATA_MINING PL/SQL package contains these two procedures:

- EXPORT_MODEL
- IMPORT_MODEL

For more information about these procedures, refer to the *Oracle Database PL/SQL Packages and Types Reference*.

Java APIs for Exporting and Importing Models

Oracle Database implements the industry-standard Java Data Mining (JDM) API Specification, which includes these two interfaces:

- `javax.datamining.task.ExportTask`
- `javax.datamining.task.ImportTask`

For more information about the standard JDM API, refer to the Java Help for the JSR-73 Specification, which is available on the Oracle Technology Network at

<http://www.oracle.com/technology/products/bi/odm/JSR-73/index.html>

Tables Created By Exporting and Importing Models

The Data Mining export and import utilities create tables in the user's schema that are for internal use only:

- DM\$P_MODEL_EXPIMP_TEMP. Used for internal purposes during export and import, and provides a job history.
- DM\$P_MODEL_IMPORT_TEMP. Used only for internal purposes during import.
- DM\$P_MODEL_TABKEY_TEMP. Used only for internal purposes during export and import.

Do not alter these tables. However, you may drop them when no export or import job is running. The utilities will re-create them for the next job.

Example: Exporting and Importing Models

This example creates a dump file containing two models and imports the models from the dump file.

This example was generated on a Linux system. The directory object MYDIR identifies the path `/scratch/dumpfiles`.

Exporting Models from the DMUSER Schema

In this example, the DMUSER schema contains two mining models.

```
SQL> CONNECT dmuser
Enter password: password

SQL> SELECT model_name FROM user_mining_models;

MODEL_NAME
-----
GLMR_SH_REGR_SAMPLE
SVMC_SH_CLAS_SAMPLE
```

The following command exports all models from DMUSER to the directory identified by MYDIR.

```
SQL> EXECUTE dbms_data_mining.export_model('all_dmuser_models.dmp', 'mydir');
```

An export or import creates a log file in the same directory as the dump file. Error messages are returned to the current output device (such as the screen), and the log file may provide additional information.

This sample export created two files in the MYDIR directory:

- A dump file named ALL_DMUSER_MODELS01.DMP (note the 2-digit suffix added to the name)
- A log file with the name DMUSER_EXP_920.LOG

For detailed information about the default names of files, see *Oracle Database PL/SQL Packages and Types Reference*.

You can view the log file using a system command or editor. You must know the path of the physical directory in order to locate the file.

DMUSER_EXP_920.LOG lists the two exported mining models and supporting objects .

```
Starting "DMUSER"."DMUSER_exp_17":  DM_EXPIMP_JOB_ID=17
Estimate in progress using BLOCKS method...
Processing object type TABLE_EXPORT/TABLE/TABLE_DATA
Total estimation using BLOCKS method: 1.062 MB
Processing object type TABLE_EXPORT/TABLE/TABLE
Processing object type TABLE_EXPORT/TABLE/GRANT/OWNER_GRANT/OBJECT_GRANT
Processing object type TABLE_EXPORT/TABLE/INDEX/INDEX
Processing object type TABLE_EXPORT/TABLE/CONSTRAINT/CONSTRAINT
Processing object type TABLE_EXPORT/TABLE/INDEX/STATISTICS/INDEX_STATISTICS
Processing object type TABLE_EXPORT/TABLE/STATISTICS/TABLE_STATISTICS
Processing object type TABLE_EXPORT/TABLE/POST_INSTANCE/PROCACT_INSTANCE
. . exported "DMUSER"."DM$PIGLMR_SH_REGR_SAMPLE"          7.085 KB      14 rows
. . exported "DMUSER"."DM$PISVMC_SH_CLAS_SAMPLE"         7.507 KB      17 rows
. . exported "DMUSER"."DM$PCGLMR_SH_REGR_SAMPLE"        53.27 KB     2278 rows
. . exported "DMUSER"."DM$PAGLMR_SH_REGR_SAMPLE"        5.796 KB      24 rows
. . exported "DMUSER"."DM$PBGLMR_SH_REGR_SAMPLE"       15.76 KB      67 rows
. . exported "DMUSER"."DM$PDGLMR_SH_REGR_SAMPLE"         8 KB         66 rows
. . exported "DMUSER"."DM$PDSVMC_SH_CLAS_SAMPLE"        9.023 KB      88 rows
. . exported "DMUSER"."DM$PFGLMR_SH_REGR_SAMPLE"        5.656 KB      10 rows
. . exported "DMUSER"."DM$POSVMC_SH_CLAS_SAMPLE"        5.320 KB       8 rows
. . exported "DMUSER"."DM$PXSVMC_SH_CLAS_SAMPLE"       7.265 KB      77 rows
. . exported "DMUSER"."DM$PZSVMC_SH_CLAS_SAMPLE"        6.164 KB       1 rows
. . exported "DMUSER"."DM$P_MODEL_EXPIMP_TEMP"         5.921 KB       2 rows
Master table "DMUSER"."DMUSER_exp_17" successfully loaded/unloaded
*****
Dump file set for DMUSER.DMUSER_exp_17 is:
  /scratch/dumpfiles/all_dmuser_models01.dmp
Job "DMUSER"."DMUSER_exp_17" successfully completed at 10:30:19
```

Importing Models Into the Same Schema

The exported models still exist in DMUSER. In this example, we drop the models before importing from the dump file. An import will not overwrite an existing model with the same name.

```
SQL> EXECUTE dbms_data_mining.drop_model('GLMR_SH_REGR_SAMPLE');
SQL> EXECUTE dbms_data_mining.drop_model('SVMC_SH_CLAS_SAMPLE');
```

The following command restores all models from the dump file to the DMUSER schema.

```
SQL> EXECUTE dbms_data_mining.import_model('all_dmuser_models01.dmp', 'mydir');
```

Importing Models Into a Different Schema

A user with the necessary privileges can load the models from a dump file into a different schema. The target schema must have the same permissions and have access to the same tablespace as the schema from which the models were exported.

```
CREATE USER dmuser2 IDENTIFIED BY password
        default tablespace dmuser
        temporary tablespace temp
        quota unlimited on dmuser;
```

```
@ $ORACLE_HOME/rdbms/demo/dmshgrants sh_password dmuser2
```

The import command, also executed as SYS, loads the two models into the DMUSER2 schema.

```
SQL> EXECUTE dbms_data_mining.import_model('all_dmuser_models01.dmp', 'mydir',
null, null, null, 'todmuser2', 'dmuser:dmuser2');
```

A parameter specifies TODMUSER2.LOG as the name of the log file; the .LOG extension is added automatically to the name. The log file shows the names of the imported models and supporting metadata.

```
Master table "SYS"."todmusr2" successfully loaded/unloaded
Starting "SYS"."todmusr4":  DM_EXPIMP_JOB_ID=21
Processing object type TABLE_EXPORT/TABLE/TABLE
Processing object type TABLE_EXPORT/TABLE/TABLE_DATA
. . imported "DMUSER2"."DM$PIGLMR_SH_REGR_SAMPLE"          7.085 KB      14 rows
. . imported "DMUSER2"."DM$PISVMC_SH_CLAS_SAMPLE"          7.507 KB      17 rows
. . imported "DMUSER2"."DM$PCGLMR_SH_REGR_SAMPLE"          53.27 KB     2278 rows
. . imported "DMUSER2"."DM$PAGLMR_SH_REGR_SAMPLE"          5.796 KB      24 rows
. . imported "DMUSER2"."DM$PBGLMR_SH_REGR_SAMPLE"          15.76 KB      67 rows
. . imported "DMUSER2"."DM$PDGLMR_SH_REGR_SAMPLE"           8 KB         66 rows
. . imported "DMUSER2"."DM$PDSVMC_SH_CLAS_SAMPLE"          9.023 KB      88 rows
. . imported "DMUSER2"."DM$PFGLMR_SH_REGR_SAMPLE"          5.656 KB      10 rows
. . imported "DMUSER2"."DM$POSVMC_SH_CLAS_SAMPLE"          5.320 KB       8 rows
. . imported "DMUSER2"."DM$PXSVMC_SH_CLAS_SAMPLE"          7.265 KB      77 rows
. . imported "DMUSER2"."DM$PZSVMC_SH_CLAS_SAMPLE"          6.164 KB       1 rows
. . imported "DMUSER2"."DM$P_MODEL_EXPIMP_TEMP"            5.921 KB       2 rows
Processing object type TABLE_EXPORT/TABLE/GRANT/OWNER_GRANT/OBJECT_GRANT
Processing object type TABLE_EXPORT/TABLE/INDEX/INDEX
Processing object type TABLE_EXPORT/TABLE/CONSTRAINT/CONSTRAINT
Processing object type TABLE_EXPORT/TABLE/INDEX/STATISTICS/INDEX_STATISTICS
Processing object type TABLE_EXPORT/TABLE/STATISTICS/TABLE_STATISTICS
Processing object type TABLE_EXPORT/TABLE/POST_INSTANCE/PROCACT_INSTANCE
Job "SYS"."todmusr2" successfully completed at 11:15:26
```

The Data Mining Sample Programs

A number of sample programs are available with Oracle Data Mining. These programs illustrate the many features of the PL/SQL and Java APIs.

The sample programs create a set of models in the database. You can examine the sample source code, which includes numerous comments, to familiarize yourself with the Oracle Data Mining APIs, and you can create your own models by modifying the samples.

Note: This chapter provides general information about the sample programs and the sample data.

Instructions for installing and running the sample programs are provided in [Chapter 1](#).

This chapter includes the following sections:

- [PL/SQL Programs](#)
- [Java Programs](#)
- [Text Mining Programs](#)
- [The Sample Data](#)

PL/SQL Programs

The PL/SQL sample programs illustrate the use of the `DBMS_DATA_MINING` package for creating models and the `DBMS_DATA_MINING_TRANSFORM` package for performing transformations on the mining data.

See Also:

- *Oracle Database PL/SQL Packages and Types Reference* for syntax of the PL/SQL API
- *Oracle Data Mining Application Developer's Guide* for information on the use of the APIs
- *Oracle Database SQL Language Reference* for syntax of the SQL functions for model scoring

PL/SQL Program Summaries

Summary descriptions of the PL/SQL sample programs are provided in [Table 7-1](#). For detailed descriptions of the sample programs, see the comments in the source code.

Table 7–1 Overview of the PL/SQL Sample Programs

Mining Function	Description
Classification	<p>The classification programs demonstrate various preprocessing techniques and perform the following steps:</p> <ul style="list-style-type: none"> ■ Build a classification model using training data ■ Display model details and settings ■ Test the model by applying the model on the test data ■ Present test metrics, such as confusion matrix, lift, and ROC ■ Apply the model on the scoring data ■ Present apply results ■ Present ranked apply results, influenced by a cost matrix <p><code>dmnbdemo.sql</code> illustrates Naive Bayes. <code>dmdtdemo.sql</code> illustrates Decision Tree. <code>dmsvcdem.sql</code> illustrates SVM classification. <code>dmglcdem</code> illustrates GLM classification (binary logistic regression)</p> <p>The <code>dmdtxv1ddemo.sql</code> program demonstrates cross-validation techniques for decision tree based-classification. With minor modifications, this program can be used to perform cross validation using other models/algorithms.</p>
Regression	<p><code>dmsvrдем.sql</code> uses different test metrics, but otherwise performs most of the same steps used in the classification programs. Selected attributes of the input data are preprocessed (normalized).</p> <p>NOTE: <code>dmsvrдем.sql</code> illustrates the new Automatic Data Preparation feature.</p> <p><code>dmglrdem.sql</code> illustrates GLM regression (multivariate linear regression)</p>
Anomaly Detection	<p><code>dmsvodem.sql</code> illustrates one-class SVM</p>
Association	<p><code>dmardemo.sql</code> builds an association model and presents frequent itemsets and association rules as output.</p>
Clustering	<p><code>dmkmdemo.sql</code> (<i>k</i>-Means) and <code>dmocdemo.sql</code> (0-Cluster) build clustering models and present cluster details, such as rules, centroid, and histogram for each cluster as output. The models are scored, and the probabilities associated with each cluster are returned as output. Selected attributes of the input data are preprocessed.</p> <p>NOTE: <code>dmkmdemo.sql</code> illustrates the new Automatic Data Preparation feature.</p>
Feature extraction	<p><code>dmnmdemo.sql</code> builds a feature extraction model and presents model details as the output. The model is scored, and each feature ID is associated with a probability. Selected attributes of the input data are preprocessed (normalized).</p>
Attribute importance	<p><code>dmaidemo.sql</code> builds an attribute importance model and presents a list of important attributes as the output of model details. Selected attributes of the input data are preprocessed (binned).</p>

Data Mining SQL Scoring Functions

Most of the PL/SQL sample programs use the Data Mining SQL functions for scoring. The Data Mining scoring functions can be used to apply models created with the PL/SQL API or with the Java API.

Note: The SQL functions for Data Mining are documented in *Oracle Database SQL Language Reference*. Information about these functions is also provided in *Oracle Data Mining Application Developer's Guide*.

The programs that demonstrate the Data Mining functions are listed in [Table 7-2](#).

Table 7-2 Data Mining SQL Functions in the Sample Programs

Program Name	Algorithm	SQL Functions Used
dmkmdemo.sql	k-Means	CLUSTER_ID, CLUSTER_PROBABILITY, CLUSTER_SET
dmocdemo.sql	O-Cluster	CLUSTER_ID
dmmmdemo.sql	NMF	FEATURE_ID, FEATURE_SET, FEATURE_VALUE
dmdtdemo.sql	Decision Tree	PREDICTION, PREDICTION_COST, PREDICTION_DETAILS, PREDICTION_SET
dmsvcdem.sql	SVM classification	PREDICTION, PREDICTION_PROBABILITY, PREDICTION_SET
dmsvodem.sql	One-Class SVM	PREDICTION, PREDICTION_PROBABILITY
dmsvrдем.sql	SVM regression	PREDICTION
dmtxtsvm.sql	Text mining	PREDICTION, PREDICTION_PROBABILITY

Java Programs

The Java demos illustrate the features of the Oracle Data Mining Java API, which implements Oracle-specific extensions to the Java Data Mining (JDM) 1.0.1.1 standard. The Java programs demonstrate data preprocessing and the basic mining functions. Additional Java samples demonstrate predictive analytics, import/export, and text mining.

See Also: *Oracle Data Mining Java API Reference* (javadoc) and the *Oracle Data Mining Application Developer's Guide* for information on the Java API.

Java Program Summaries

Summary descriptions of the Java sample programs are provided in [Table 7-3](#). For detailed descriptions, see the comments in the source code.

Table 7–3 Overview of the Java Sample Programs

Mining Function or Task	Description
Classification	<p>The classification programs demonstrate various preprocessing techniques and perform the following steps:</p> <ul style="list-style-type: none"> ■ Build a classification model using training data ■ Display model details and settings ■ Test the model by applying the model on the test data ■ Present test metrics, such as confusion matrix, lift, and ROC ■ Apply the model on the scoring data ■ Present apply results ■ Present ranked apply results, influenced by a cost matrix <p>The <code>dmapplydemo.java</code> program demonstrates several ways of applying a Naive Bayes model.</p> <p><code>dmglcdemo.java</code> illustrates GLM classification (binary logistic regression)</p>
Regression	<p><code>dmsvrdemo.java</code> uses different test metrics, but otherwise performs most of the same steps used in the classification programs. Selected attributes of the input data are preprocessed (normalized).</p> <p><code>dmglrdemo.java</code> illustrates GLM regression (multivariate linear regression)</p>
Association	<p><code>dmardemo.java</code> builds an association model and presents frequent itemsets and association rules as output. Selected attributes of the input data are preprocessed (binned).</p>
Clustering	<p><code>dmkmdemo.java</code> (<i>k</i>-Means) and <code>dmocdemo.java</code> (0-Cluster) build clustering models and present cluster details, such as rules, centroid, and histogram for each cluster as output. The models are scored, and the probabilities associated with each cluster are returned as output. Selected attributes of the input data are preprocessed (normalized).</p>
Feature extraction	<p><code>dmnmdemo.java</code> builds a feature extraction model and presents model details as the output. The model is scored, and each feature ID is associated with a probability. Selected attributes of the input data are preprocessed (normalized).</p>
Attribute importance	<p><code>dmaidemo.java</code> builds an attribute importance model and presents a list of important attributes as the output of model details. Selected attributes of the input data are preprocessed (binned).</p>
Data transformations	<p><code>dmxfdemo.java</code> demonstrates binning, clipping, and normalization transformations.</p>
Predictive Analytics	<p><code>dmpademo.java</code> demonstrates PREDICT, EXPLAIN, and PROFILE functions.</p>
Model import/export	<p><code>dmexpimpdemo.java</code> builds a Naive Bayes model, exports it to a dump file, then imports it from the dump file.</p>

Text Mining Programs

Oracle Data Mining can mine text columns that have undergone pre-processing by Oracle Text routines.

Oracle Text is a technology for building text query and document classification applications. It provides indexing, word and theme searching, and viewing

capabilities for text. Oracle Text is included in a general installation of Oracle Database Enterprise Edition, and therefore is already present in a database installed according to the instructions in [Chapter 1](#).

The pre-processing steps for text mining create nested table columns of type `DM_NESTED_NUMERICALS` from columns of type `VARCHAR2` or `CLOB`. Each row of the nested table specifies an attribute name and a value. The type definition is as follows.

```
CREATE OR REPLACE TYPE dm_nested_numerical AS OBJECT
  (attribute_name VARCHAR2(4000),
   value          NUMBER)
/
CREATE OR REPLACE TYPE dm_nested_numericals AS TABLE OF dm_nested_numerical
```

Terms extracted from text documents into nested tables can become generic attributes in training or scoring data.

Sample text mining programs in both PL/SQL and Java illustrate classification and feature extraction of a pre-processed text column.

Text Mining in PL/SQL

Three PL/SQL sample programs illustrate the process of text mining. One program illustrates the pre-processing that is required to prepare the data for mining. The other two programs build models that use the transformed text.

Text Transformation Demo

To prepare a column for text mining using the PL/SQL API, you must use Oracle Text routines to perform the following general steps:

1. Create a domain index on the column.
2. Use the index to extract terms from the column to a temporary table.
3. Populate a column of type `DM_NESTED_NUMERICALS` with the terms in the temporary table.

The process of term extraction using Oracle Text is illustrated in the sample program `dmtxtfe.sql`. The source code contains extensive comments that explain the steps involved in transforming text into a set of features that can be mined using Oracle Data Mining.

More details about text transformation are provided in the *Oracle Data Mining Application Developer's Guide*.

Text Transformation for the PL/SQL Text Mining Sample Programs

The `dmsb.sql` script performs the text transformation required by the PL/SQL text mining demos. There are two such sample programs: `dmtxtnmf.sql`, which builds a feature extraction model using Non-Negative Matrix Factorization, and `dmtxtsvm.sql`, which builds a classification model using Support Vector Machine. Both of these programs use the following tables, which have a nested table column of comment data:

```
MINING_BUILD_NESTED_TEXT
MINING_TEST_NESTED_TEXT
MINING_APPLY_NESTED_TEXT
```

The Sample Text Mining Models (PL/SQL)

You can run the PL/SQL text mining demo programs, `dmtxtnmf.sql` and `dmtxsvm.sql`, like the other PL/SQL programs. The models created by these programs are listed in the following example.

```
SQL> @ %ORACLE_HOME%\rdbms\demo\dmtxtnmf.sql
SQL> @ %ORACLE_HOME%\rdbms\demo\dmtxsvm.sql
SQL> SELECT model_name, mining_function, algorithm
        FROM user_mining_models;
```

MODEL_NAME	MINING_FUNCTION	ALGORITHM
T_NMF_SAMPLE	FEATURE_EXTRACTION	NONNEGATIVE_MATRIX_FACTOR
T_SVM_CLAS_SAMPLE	CLASSIFICATION	SUPPORT_VECTOR_MACHINES

Text Mining in Java

Two Java sample programs illustrate the process of text mining. One builds a feature extraction model, the other builds a classification model.

Text Transformation for the Sample Java Text Mining Programs

The Oracle Data Mining Java API provides an interface that handles the term extraction process. If you are developing data mining applications in Java, you do not need to use Oracle Text directly. However, you must ensure that Oracle Text is present in the database.

The `OraTextTransform` interface is used to perform text transformation within the Java text mining demos. There are two such sample programs: `dmtxtnmfdemo.java`, which builds a feature extraction model using Non-Negative Matrix Factorization, and `dmtxsvmdemo.java`, which builds a classification model using Support Vector Machine. Both of these programs create build, test, and apply data sets from the following tables, which have a text column of comment data:

```
MINING_BUILD_TEXT
MINING_TEST_TEXT
MINING_APPLY_TEXT
```

The Sample Text Mining Models (Java)

You can run the Java text mining sample programs, `dmtxtnmfdemo.java` and `dmtxsvmdemo.java`, like the other Java programs. The models created by these programs are shown in the following example.

```
java dmtxtnmfdemo host:port:SID dmuser password
java dmtxsvmdemo host:port:SID dmuser password
SQLPLUS dmuser
Enter password: password
SQL> SELECT model_name, mining_function, algorithm
        FROM user_mining_models;
```

MODEL_NAME	MINING_FUNCTION	ALGORITHM
txtnmfModel_jdm	FEATURE_EXTRACTION	NONNEGATIVE_MATRIX_FACTOR
txsvmModel_jdm	CLASSIFICATION	SUPPORT_VECTOR_MACHINES

The Sample Data

The `dmsh.sql` script creates views, tables, and indexes in the user's schema. The views define columns of customer data from tables in the `SH` schema. This data is used by the Data Mining sample programs. The tables reference the same columns in `SH`, but they include an extra `COMMENTS` column for text mining. The indexes are used to extract terms from the text in the `COMMENTS` column and build a nested table column.

Customer Data for Data Mining

Views in the data mining user's schema define columns of data from the `CUSTOMERS`, `SALES`, `PRODUCTS`, `COUNTRIES`, and `SUPPLEMENTARY_DEMOGRAPHICS` tables in the `SH` schema. You can list these views with the following SQL statements.

```
SQL>CONNECT dmuser
Enter password: password
SQL>SELECT view_name FROM user_views;
```

The views are listed in [Table 7-4](#).

Table 7-4 Views Used by the Data Mining Sample Programs

View Name	Description
<code>MINING_DATA_APPLY_STR_V</code>	Scoring data for o-cluster
<code>MINING_DATA_BUILD_STR_V</code>	Training data for o-cluster
<code>MINING_DATA_APPLY_V</code>	Scoring data for data mining (not text mining)
<code>MINING_DATA_BUILD_V</code>	Training data for data mining (not text mining)
<code>MINING_DATA_TEST_V</code>	Test data for data mining (not text mining)
<code>MARKET_BASKET_V</code>	Data for association rules
<code>MINING_DATA_ONE_CLASS_V</code>	Data for one-class SVM

You can see the references to tables in `SH` by listing the view definitions. The definition of the view `MINING_DATA_BUILD_V` is shown as follows.

```
SQL> set long 1000000
SQL> set longc 100000
SQL> set pagesize 100
SQL> SELECT text FROM all_views WHERE
       owner='dmuser3'AND view_name='mining_data_build_v';

       SELECT a.CUST_ID, a.CUST_GENDER, 2003-a.CUST_YEAR_OF_BIRTH AGE,
              a.CUST_MARITAL_STATUS, c.COUNTRY_NAME, a.CUST_INCOME_LEVEL,
              b.EDUCATION, b.OCCUPATION, b.HOUSEHOLD_SIZE, b.YRS_RESIDENCE,
              b.AFFINITY_CARD, b.BULK_PACK_DISKETTES, b.FLAT_PANEL_MONITOR,
              b.HOME_THEATER_PACKAGE, b.BOOKKEEPING_APPLICATION,
              b.PRINTER_SUPPLIES, b.Y_BOX_GAMES, b.OS_DOC_SET_KANJI
       FROM   sh.customers a,
              sh.supplementary_demographics b,
              sh.countries c
       WHERE  a.CUST_ID = b.CUST_ID AND a.country_id = c.country_id
              AND a.cust_id between 101501 and 103000
```

The views are used to build, test, and score the sample models. Each view has a `CUSTOMER_ID` column, which is the case ID, and an `AFFINITY_CARD` column, which is the target used by the predictive models. Most of the views provide data for 1500

customers (1500 rows). The view used by the One-Class SVM model has data for 940 customers.

The columns of training data in the `MINING_DATA_BUILD_V` view are listed in the following example.

```
SQL> DESCRIBE mining_data_build_v
```

<code>CUST_ID</code>	<code>NOT NULL</code>	<code>NUMBER</code>
<code>CUST_GENDER</code>	<code>NOT NULL</code>	<code>CHAR(1)</code>
<code>AGE</code>		<code>NUMBER</code>
<code>CUST_MARITAL_STATUS</code>		<code>VARCHAR2(20)</code>
<code>COUNTRY_NAME</code>	<code>NOT NULL</code>	<code>VARCHAR2(40)</code>
<code>CUST_INCOME_LEVEL</code>		<code>VARCHAR2(30)</code>
<code>EDUCATION</code>		<code>VARCHAR2(21)</code>
<code>OCCUPATION</code>		<code>VARCHAR2(21)</code>
<code>HOUSEHOLD_SIZE</code>		<code>VARCHAR2(21)</code>
<code>YRS_RESIDENCE</code>		<code>NUMBER</code>
<code>AFFINITY_CARD</code>		<code>NUMBER(10)</code>
<code>BULK_PACK_DISKETTES</code>		<code>NUMBER(10)</code>
<code>FLAT_PANEL_MONITOR</code>		<code>NUMBER(10)</code>
<code>HOME_THEATER_PACKAGE</code>		<code>NUMBER(10)</code>
<code>BOOKKEEPING_APPLICATION</code>		<code>NUMBER(10)</code>
<code>PRINTER_SUPPLIES</code>		<code>NUMBER(10)</code>
<code>Y_BOX_GAMES</code>		<code>NUMBER(10)</code>
<code>OS_DOC_SET_KANJI</code>		<code>NUMBER(10)</code>

Market Basket Data for Association Rules

The association demos use the `MARKET_BASKET_V` data set, which includes columns of products from the `PRODUCTS` table and the `CUSTOMER_ID` column from the `CUSTOMERS` table in `SH`. The columns of the `MARKET_BASKET_V` view are listed in the following example.

```
SQL> DESCRIBE market_basket_v
```

<code>CUST_ID</code>	<code>NOT NULL</code>	<code>NUMBER</code>
<code>EXTENSION_CABLE</code>		<code>NUMBER</code>
<code>FLAT_PANEL_MONITOR</code>		<code>NUMBER</code>
<code>CD_RW_HIGH_SPEED_5_PACK</code>		<code>NUMBER</code>
<code>ENVOY_256MB_40GB</code>		<code>NUMBER</code>
<code>ENVOY_AMBASSADOR</code>		<code>NUMBER</code>
<code>EXTERNAL_8X_CD_ROM</code>		<code>NUMBER</code>
<code>KEYBOARD_WRIST_REST</code>		<code>NUMBER</code>
<code>SM26273_BLACK_INK_CARTRIDGE</code>		<code>NUMBER</code>
<code>MOUSE_PAD</code>		<code>NUMBER</code>
<code>MULTIMEDIA_SPEAKERS_3INCH</code>		<code>NUMBER</code>
<code>OS_DOC_SET_ENGLISH</code>		<code>NUMBER</code>
<code>SIMM_16MB_PCMCIAII_CARD</code>		<code>NUMBER</code>
<code>STANDARD_MOUSE</code>		<code>NUMBER</code>

Customer Data for Text Mining

The text mining demos use the same customer data from tables in `SH`, but they include either an extra text column or a collection type column. The collection type is a nested table of type `DM_NESTED_NUMERICALS`.

You can list these tables with the following SQL statements.

```
SQL>CONNECT dmuser
Enter password: password
```

```
SQL>SELECT table_name FROM user_tables WHERE table_name LIKE '%MINING%';
```

The text mining tables are listed in [Table 7-5](#).

Table 7-5 Tables Used by the Text Mining Sample Programs

Table Name	Description
MINING_APPLY_NESTED_TEXT	Apply table with COMMENTS column as DM_NESTED_NUMERICALS
MINING_BUILD_NESTED_TEXT	Build table with COMMENTS column as DM_NESTED_NUMERICALS
MINING_TEST_NESTED_TEXT	Test table with COMMENTS column as DM_NESTED_NUMERICALS
MINING_APPLY_TEXT	Apply table with COMMENTS column as VARCHAR2 (4000)
MINING_BUILD_TEXT	Build table with COMMENTS column as VARCHAR2 (4000)
MINING_TEST_TEXT	Test table with COMMENTS column as VARCHAR2 (4000)

In the MINING_BUILD_TEXT, MINING_TEST_TEXT, and MINING_APPLY_TEXT tables, the COMMENTS column is of type VARCHAR2 (4000).

```
SQL> DESCRIBE mining_build_text
```

Name	Null?	Type
CUST_ID	NOT NULL	NUMBER
CUST_GENDER	NOT NULL	CHAR(1)
AGE		NUMBER
CUST_MARITAL_STATUS		VARCHAR2(20)
COUNTRY_NAME	NOT NULL	VARCHAR2(40)
CUST_INCOME_LEVEL		VARCHAR2(30)
EDUCATION		VARCHAR2(21)
OCCUPATION		VARCHAR2(21)
HOUSEHOLD_SIZE		VARCHAR2(21)
YRS_RESIDENCE		NUMBER
AFFINITY_CARD		NUMBER(10)
BULK_PACK_DISKETTES		NUMBER(10)
FLAT_PANEL_MONITOR		NUMBER(10)
HOME_THEATER_PACKAGE		NUMBER(10)
BOOKKEEPING_APPLICATION		NUMBER(10)
PRINTER_SUPPLIES		NUMBER(10)
Y_BOX_GAMES		NUMBER(10)
OS_DOC_SET_KANJI		NUMBER(10)
COMMENTS		VARCHAR2(4000)

In the MINING_*_NESTED_TEXT tables, the COMMENTS column is of type DM_NESTED_NUMERICALS.

```
SQL> DESCRIBE mining_build_nested_text
```

Name	Null?	Type
CUST_ID	NOT NULL	NUMBER
CUST_GENDER	NOT NULL	CHAR(1)
AGE		NUMBER
CUST_MARITAL_STATUS		VARCHAR2(20)
COUNTRY_NAME	NOT NULL	VARCHAR2(40)
CUST_INCOME_LEVEL		VARCHAR2(30)
EDUCATION		VARCHAR2(21)
OCCUPATION		VARCHAR2(21)
HOUSEHOLD_SIZE		VARCHAR2(21)
YRS_RESIDENCE		NUMBER
AFFINITY_CARD		NUMBER(10)

BULK_PACK_DISKETTES	NUMBER(10)
FLAT_PANEL_MONITOR	NUMBER(10)
HOME_THEATER_PACKAGE	NUMBER(10)
BOOKKEEPING_APPLICATION	NUMBER(10)
PRINTER_SUPPLIES	NUMBER(10)
Y_BOX_GAMES	NUMBER(10)
OS_DOC_SET_KANJI	NUMBER(10)
COMMENTS	DM_NESTED_NUMERICALS

The process of extracting terms from a text column into a nested table column is described in ["Text Mining Programs"](#) on page 7-4 and in *Oracle Data Mining Application Developer's Guide*.

Index

A

Adaptive Bayes Network, xiii
add-in
 See Oracle Spreadsheet Add-In for Predictive Analytics
ADP
 See Automatic Data Preparation
ALL_MINING_MODEL_ATTRIBUTES, xi, xiii, 6-1, 6-2
ALL_MINING_MODEL_SETTINGS, xi, xiii, 6-1, 6-3
ALL_MINING_MODELS, xi, xiii, 6-1
ALL_OBJECTS, xii
ALTER ANY MINING MODEL, 4-2
ALTER MINING MODEL, 4-3
anomaly detection, 1-11
Apriori, 1-11, 1-12
association rules, 1-11, 7-8
attribute importance, 1-11
attributes, 6-2
AUDIT, xii, 4-2, 6-4, 6-5
AUDIT ANY, 4-2
AUDIT_TRAIL parameter, 6-4
Automatic Data Preparation, 7-2

B

batch scoring, viii
binning, 1-12
BLAST, xiii

C

catalog views, xi, 6-1
classification, 1-11
CLASSPATH, 1-13
clipping, 1-12
CLUSTER_ID, 7-3
CLUSTER_PROBABILITY, 7-3
CLUSTER_SET, 7-3
clustering, 1-11
COMMENT, 4-2, 6-2, 6-4
COMMENT ANY MINING MODEL, 4-2
COMMENT ANY MODEL, 6-3
cost matrix, 4-3
CREATE ANY DIRECTORY, 6-6

CREATE ANY MINING MODEL, 4-2
CREATE MINING MODEL, 4-1, 4-2

D

data dictionary views, xi, 6-1
Data Mining demo user, 1-10, 2-1, 2-3, 4-1, 5-1
Data Mining option, 1-1
Data Mining user, 2-1, 2-3, 5-1
Database Companion, 1-7
database tuning, viii
DBA_AUDIT_OBJECT, 6-5
DBA_MINING_MODEL_ATTRIBUTES, 6-1
DBA_MINING_MODEL_SETTINGS, 6-1
DBA_MINING_MODELS, 6-1
DBA_OBJ_AUDIT_OPTS, 6-5
DBA_OBJECTS, xii
DBMS_DATA_MINING, 7-1
DBMS_DATA_MINING_TRANSFORM, 7-1
Decision Tree, 1-11, 1-12
demos
 See sample programs
desupported and deprecated features, xii
directory objects, 6-6
DM\$P_MODEL_EXPIMP_TEMP table, 6-7
DM\$P_MODEL_IMPORT_TEMP table, 6-7
DM\$P_MODEL_TABKEY_TEMP table, 6-7
DM_NESTED_NUMERICALS, 7-5
DM_USER_MODELS, xiii
DMEIDMSYS, 3-2
DMSH, 1-11, 2-3, 7-5, 7-7
DMSHGRANTS, 1-11, 2-3
DMSYS, xi, xii, 3-1, 3-2
downgrading, 3-1, 3-3
DROP ANY MINING MODEL, 4-2

E

Enterprise Edition, 1-2
EXPDP, 6-5
EXPLAIN, 5-1
EXPORT_MODEL, 6-7
exporting, 1-12, 3-2, 4-2, 6-5, 6-7

F

feature extraction, 1-11
FEATURE_ID, 7-3
FEATURE_SET, 7-3
FEATURE_VALUE, 7-3

G

Generalized Linear Models, 1-11, 7-2, 7-4
GET_DEFAULT_SETTINGS, xiii
GET_MODEL_DETAILS, 6-2
GET_MODEL_SETTINGS, xiii
GET_MODEL_SIGNATURE, xiii
GLM, 1-11
 See Generalized Linear Models

I

IMPDP, 6-5
IMPORT_MODEL, 6-7
importing, 1-12, 3-2, 4-2, 6-5, 6-8
INIT.ORA, viii
installation
 Database Companion, 1-7
 Oracle Client, 2-2

J

Java API, 1-12, 6-7, 7-3
javax.datamining.task.ExportTask interface, 6-7
javax.datamining.task.ImportTask interface, 6-7
JDM 1.0 standard, 6-7, 7-3
JSR-73 specification, 6-7, 7-3

K

k-Means, 1-11, 1-12, 7-3

L

linear regression, 1-11, 1-12, 7-2
local database, 2-1
logistic regression, 1-11, 1-12, 7-2

M

memory tuning parameters, viii
Minimum Description Length, 1-12
mining models
 about, 6-1
 adding a comment, 6-3
 applying, 4-3
 auditing, 6-4
 changing the name, 4-3
 listing, 1-13, 6-1
 settings, 6-3
 upgrading, 3-1
 viewing, 4-3
model details, 6-2
model signature, xiii, 6-2

N

Naive Bayes, 1-11, 1-12
nested tables, xii, 7-5
new features, xi
NMF
 See Non-Negative Matrix Factorization
Non-Negative Matrix Factorization, 1-11, 1-12, 7-3
normalization, 1-12

O

object privileges, 4-3
O-Cluster, 1-11, 1-12, 7-3
One-Class SVM, 1-11, 7-3
Oracle Client, 5-1
Oracle Data Pump, 6-5
Oracle Database Upgrade Assistant, 3-2
Oracle Net Configuration Assistant, 2-3
Oracle Net Service Name, 2-2, 2-3, 5-2
Oracle Objects for OLE, 5-1
Oracle Spreadsheet Add-In for Predictive Analytics, 5-1
Oracle Text, 1-11, 7-4
Oracle Universal Installer, 1-2
OraTextTransform interface, 7-6

P

parallel execution, viii
performance, viii
PGA_AGGREGATE_TARGET, viii
PL/SQL API, 1-11
PL/SQL sample programs, 7-1
PREDICT, 5-1
PREDICTION, 7-3
PREDICTION_COST, 7-3
PREDICTION_DETAILS, 7-3
PREDICTION_PROBABILITY, 7-3
PREDICTION_SET, 7-3
predictive analytics, 1-12, 5-1
privileges for data mining, xii, 1-10, 2-1, 4-1, 5-1, 6-6
PROFILE, 5-1

R

real-time scoring, viii
regression, 1-11
remote database, 2-2

S

sample programs, 1-7, 2-3, 7-1, 7-3
 data used by, 7-7
 Java, 1-12, 1-13, 7-3
 PL/SQL, 1-11, 1-12, 7-1
sample schemas, 1-1, 2-3
scoping, xii
scoring, viii, 7-2
Scoring Engine, xii
security, xii, 4-1

SELECT ANY MINING MODEL, 4-2
SELECT MINING MODEL, 4-3
settings, 6-3
SGA_TARGET, viii
SH schema, 2-3, 7-7
shared cursor, viii
sparse data, xii
Spreadsheet Add-In
 See Oracle Spreadsheet Add-In for Predictive
 Analytics
SQL AUDIT, 6-4
SQL COMMENT, 6-3
SQL Data Mining functions, 7-2
SQL*Plus, 2-1
Support Vector Machine, 1-11, 1-12, 7-3
SVM
 See Also One-Class SVM
 See Support Vector Machine
SYS, xi, 3-1, 3-2
system privileges, 4-2

T

text mining, 1-11, 1-12, 7-3, 7-4, 7-8
transformations, 1-12
tuning, viii

U

upgrading, 3-1
user ID, 1-10, 2-1, 2-3, 4-1, 5-1, 6-6
USER_AUDIT_OBJECT, 6-5
USER_MINING_MODEL_ATTRIBUTES, 6-1
USER_MINING_MODEL_SETTINGS, 6-1
USER_MINING_MODELS, 1-13, 6-1
USER_OBJ_AUDIT_OPTS, 6-5
USER_OBJECTS, xii
users, 6-6

