Лабораторна робота 1. Setting up Oracle Data Miner 11g Release 2 by Using SQL Developer 3.1

***What is Data Mining?***

***Many definitions, here is one:***

***Data mining is the process of automatically discovering useful information in large data repository.***

**Setting up Oracle Data Miner 11g Release 2 by Using SQL Developer 3.1**

**Oracle Data Mining (ODM),** is a component of the Oracle Advanced Analytics Option for Oracle Database 11g Enterprise Edition. ODM provides a collection of in-database data mining algorithms that solve a wide range of business problems. It provides the necessary mining model building, testing and scoring capabilities. Because of data, models, and results remain in Oracle Database, data movement is eliminated, security is maximized, and information latency is minimized.

The purpose of this blog is to explain various steps involved in setting up Oracle Data Miner 11g Release 2 by using SQL Developer 3.1.

**Pre-requisites for Setting up Oracle Data Miner**

The following configuration elements and information is required for setting up Oracle Data Miner.

**Oracle Database 11g Enterprise Edition, Release 2** (11.2.0.1) with following services which are critical in support of Oracle Data Miner

* Oracle Data Mining- ODM provides the necessary mining model building, testing and scoring capabilities
* Oracle XML DB- Provides services to manage the Data Miner Repository metadata, such as the details of the workflow specifications
* Oracle Scheduler- Provides the engine of mining the Data Miner workflows
* Oracle Text - Provides services necessary to support text mining

**Oracle SQL Developer Version 3.1 (or Later)**

**Database Information for SYS user to perform the configuration work for the setup –** such as password, database port number, SID value or Service Name

**Steps to setup the Oracle Data Miner**

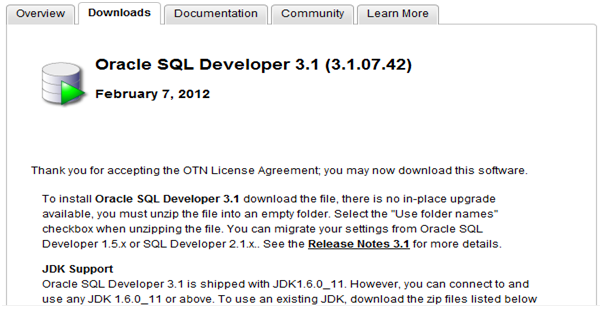
For setting up the Oracle Data Miner, user needs to perform following steps

* SQL Developer 3.1 ( or Later) Install
* Set up Data Miner by using SQL Developer

**Installing SQL Developer**

Oracle SQL Developer can be downloaded from Oracle Technology Network (OTN), it is freeware. To download Oracle SQL Developer 3.1 (or later), click on following link, after accepting agreement, choose SQL Developer download for appropriate operating system.

<http://www.oracle.com/technetwork/developer-tools/sql-developer/downloads/index.html>

[](http://2.bp.blogspot.com/-yMhtUvx9jbI/T-115eHeEwI/AAAAAAAABEo/7QvHqH0IPSM/s1600/pic1.png)

After downloading file, unzip the archive in the empty folder. There is no formal installation routine and there is no in-pace upgrade available for SQL Developer.

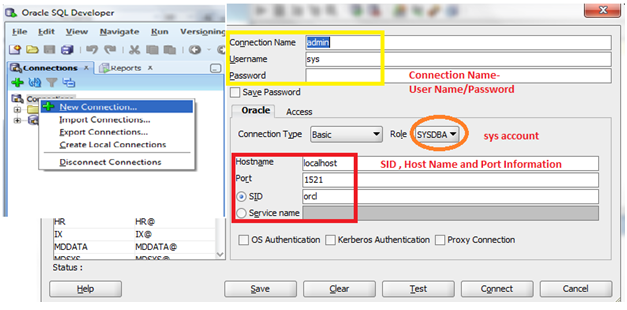
Oracle SQL Developer can be downloaded from Oracle Technology Network (OTN), it is freeware. To download Oracle SQL Developer 3.1 (or later), click on following link, after accepting agreement, choose SQL Developer download for chosen operating system.

**Setting up Data Miner**

**Create SQL Developer Connection for SYS**

This connection will be used to create the Data Miner user and install the Oracle Data Miner repository.

**Create a database user account for data mining**

[](http://1.bp.blogspot.com/-96GStbe491U/T-14RJ3nnhI/AAAAAAAABEw/37qW1rGL1IQ/s1600/pic2.png)

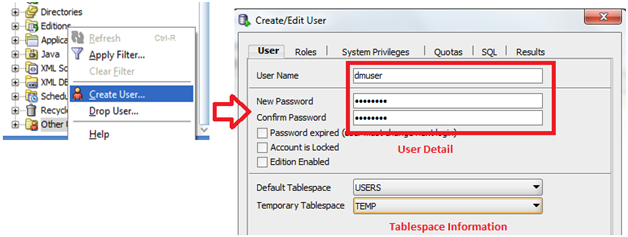
Create a database account for data mining user, the user information as follows

Name of User – dmuser

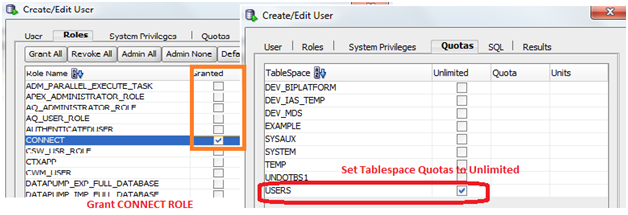
Password – XXX

Default Tablespace –USERS

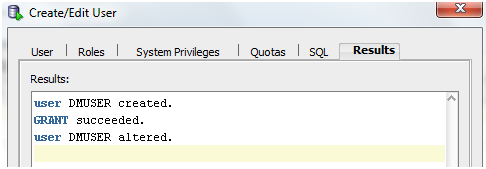
Temporary Tablespace –TEMP

[](http://4.bp.blogspot.com/-xE4rcp-OZTo/T-14nN2j6xI/AAAAAAAABE4/qcUoIwQzGto/s1600/pic3.png)

Grant CONNECT Role to dmuser and then set Tablespace Quota for USERS Tablespace to Unlimited as shown, then click on **APPLY** to create dmuser with all selected parameters.

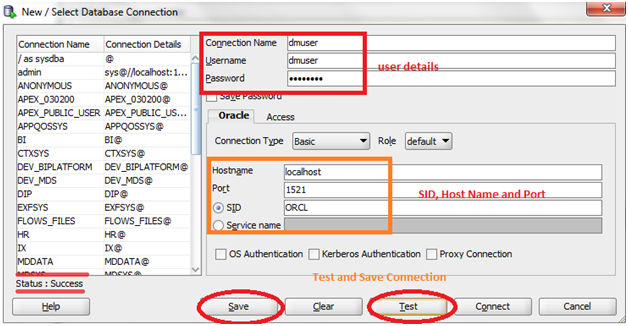
[](http://1.bp.blogspot.com/-6KT3_ydgKoQ/T-14_MMweYI/AAAAAAAABFA/HaucOE0I91c/s1600/pic4.png)

Creation of user confirmed.

[](http://2.bp.blogspot.com/-P4LKfHTexko/T-15U0v2vMI/AAAAAAAABFI/0uOSUTDpkEY/s1600/pic5.png)

**Create a SQL Developer connection for the data mining user**

Create connection for Data Miner user (dmuser)

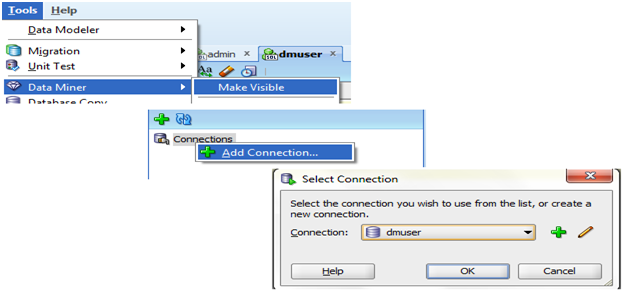
[](http://4.bp.blogspot.com/-Fw5ySp1hFeU/T-15oehY0TI/AAAAAAAABFQ/LhzsQQLkxSs/s1600/pic6.png)

**Enable the Data Miner GUI and user**

Display data miner tab similar to the normal connection tab by selecting **Tool > Data Miner > Make Visible**

On data miner tab, click on Add Connection and then select the dmuser (data mining user) from the list of connections available.

The result of execution of these steps a data miner user should appear under the Connection node in the Data Miner Tab.

[](http://4.bp.blogspot.com/-gyKlHho4Bw0/T-15-TG_jsI/AAAAAAAABFY/mY-u0mUJIkk/s1600/pic7.png)

**Install the Oracle Data Miner Repository**

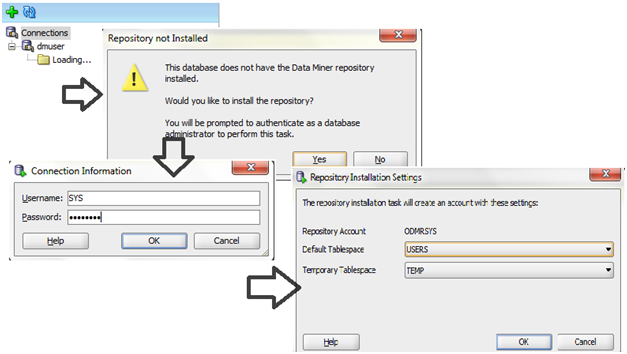
The following automated process is used to create the data miner repository

Double-click the data minder user on the Data Miner tab

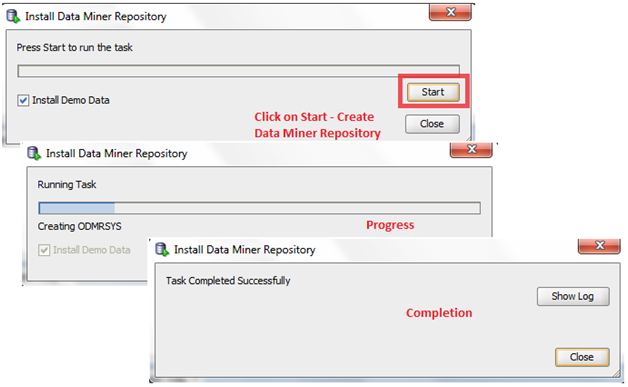
A message tells you that the Data Miner Repository is not installed in the database and asks whether you want to install the repository, click ‘Yes’ to launch the installation.

Enter the password for SYS user and click ‘OK’ to continue

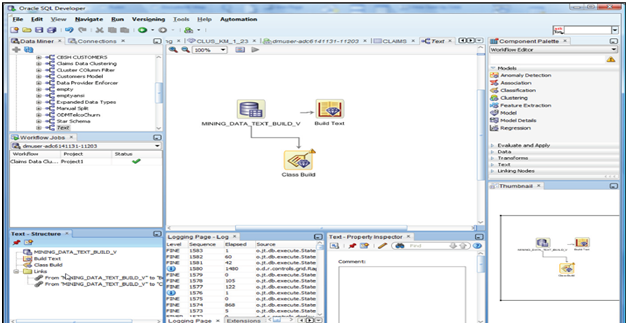
Select the Default and Temporary table spaces for the ODMSYS user, this user is created as part of the repository installation process. Click ‘OK’ to continue.

[](http://1.bp.blogspot.com/-Gl9-y69Ywsw/T-16dBmXePI/AAAAAAAABFg/t3qKfrwebAA/s1600/pic8.png)

**Start Install Data Mining Repository**

[](http://3.bp.blogspot.com/-OjbaupUkZ3c/T-161rB0IRI/AAAAAAAABFo/zg4HnqKfk2U/s1600/pic9.png)

Once repository installation completion confirms that the Oracle Data Miner setup is finished. The data miner interface is ready for use.

[](http://1.bp.blogspot.com/-8qZs6dMOGsk/T-17JQicaBI/AAAAAAAABFw/6IO74noRooo/s1600/pic10.png)

# Oracle data mining tutorial, data mining techniques: classification

Run “Oracle data mining tutorial, data mining techniques\_ classification\_(360p)” file and listen

**-----------------------------------------------------------------------------------------------------------------------------------------------------------------**

**Data Mining Using the RDOM Package**

*By Casimir Saternos*

**A primer on using the open-source R statistical analysis language with Oracle Database Enterprise Edition**

Published July 2010  
 **Downloads:**

* [Oracle Database 11g Enterprise Edition](http://www.oracle.com/technetwork/database/enterprise-edition/downloads/index.html)

Data mining, predictive analysis, and statistical techniques generally do not make headlines. However, in the last year, the New York Times cited Google’s Chief Economist Hal Varian as [predicting a significant demand for statisticians](http://www.nytimes.com/2009/08/06/technology/06stats.html?_r=2) in the next ten years. They also posted an article about the [growing popularity of the R language](http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html) as a tool for data analysis.

[R](http://www.r-project.org/) is an open source alternative to statistical software such as SPSS and SAS. It is similar to the S language (developed by John Chambers and others at Bell Laboratories) and is also influenced by Scheme (a major Lisp dialect). It can be used to perform data analysis and visualization through an interactive programming environment.

The Oracle Database includes extensive statistical and analytical functionality. Many features, including Oracle SQL analytic functions and the [DBMS\_STAT\_FUNCS](http://www.oracle.com/technology/products/bi/stats_fns/index.html) package, are available in standard installations. Oracle Database also has optional packages that complement these capabilities, including [Oracle OLAP](http://www.oracle.com/technetwork/database/options/olap/index.html).

Perhaps the most powerful of all advanced analytical capabilities is the [Oracle Data Mining](http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/overview/index.html) (ODM) option, which provides 12 data mining algorithms for discovering patterns and relationships and building and applying descriptive and predictive data mining models inside the database.

This article will demonstrate the use of [Oracle Database 11g Enterprise Edition](http://www.oracle.com/technetwork/database/enterprise-edition/overview/index.html) with the ODM option accessed by an R interface (RODM). ODM allows for analysis of data within the database. This eliminates the time, expense, and resources required to transfer data. The data remains secure in the database resulting in fewer problems due to information latency.

Oracle Database also has a graphical user interface to ODM called [Oracle Data Miner](http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/overview/index-100612.html) “Classic” and a totally new [Oracle Data Miner 11g Release 2](http://blogs.oracle.com/datamining/2010/02/get_ready_for_the_new_oracle_data_miner_11gr2_gui_1.html) workflow GUI, an extension to SQL Developer, that will be available to early adopters soon. And now, for data analysts who are already familiar with the open source R language, there is now another solution: the RODM package.

The [RODM](http://cran.fhcrc.org/web/packages/RODM/index.html) package allows R users to interact with the Oracle Database and ODM functionality. Users can analyze and manipulate data without the use of SQL or PL/SQL. Written by Pablo Tamayo and Ari Mozes, it is available for [download](http://cran.fhcrc.org/web/packages/RODM/index.html) from the Comprehensive R Archive Network (CRAN). The package depends upon the RODBC package to make Oracle Database connections and do basic data manipulation.

RODM allows R users to access the power of the ODM in-database functions using the familiar R syntax. RODM provides a powerful environment for prototyping data analysis and data mining methodologies. RODM is especially useful for:

* Quick prototyping of vertical or domain-based applications where the Oracle Database supports the application
* Scripting of "production" data mining methodologies
* Customizing graphics of ODM data mining results (examples: classification, regression, anomaly detection)

The RODM interface allows R users to mine data using ODM from the R programming environment. It consists of a set of function wrappers written in source R language that pass data and parameters from the R environment to Oracle Database 11g Enterprise Edition as standard user PL/SQL queries via an Open Database Connectivity (ODBC) interface. The RODM interface code is a thin layer of logic and SQL that calls through an ODBC interface. RODM does not use or expose any Oracle product code because it is completely an external interface and not part of any Oracle product. RODM is similar to the example scripts (for example, the PL/SQL demo code) that illustrate the use of ODM, for example, how to create data mining models, pass arguments, retrieve results, and so on.

### Software Prerequisites

If you have not done so already, install the R language on your Microsoft Windows workstation. You can then install any R packages that you would like, including the RODM package.

Please note that you will need to access (or install) an Oracle 11g database with the ODM option to use RODM. If you don't have an installed Oracle database in place and you need to install one from scratch, we strongly recommend that you follow the guidelines in the Oracle Data Mining Administrator's Guide. Oracle Database Express Edition (Oracle Database XE) does not include this option. You will also need a database user with appropriate security and privileges as well as an ODBC connection to the database. This database user should have privileges to connect to the database, create tables, create views, and create mining models.

The following is a script to create such a user. This script (modified for your environment) would need to be run by a DBA or other person with administrative privileges.

## Introduction to R

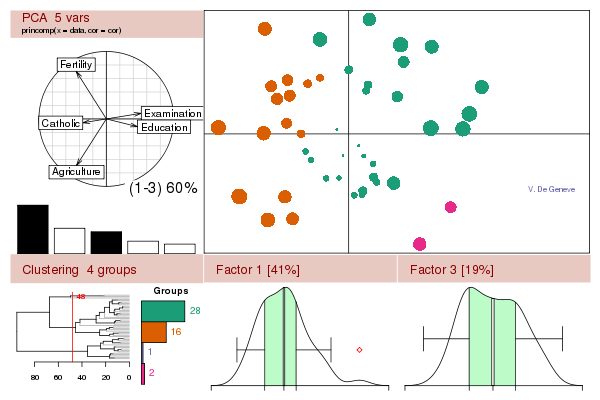
R is a language and environment for statistical computing and graphics. It is a [GNU project](http://www.gnu.org) which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R.

R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.

One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control.

R is available as Free Software under the terms of the [Free Software Foundation](http://www.gnu.org)'s [GNU General Public License](http://www.r-project.org/COPYING) in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS.

# The R Project for Statistical Computing

[](http://www.r-project.org/misc/acpclust.R)

Go to [3] and download free software environment for statistical computing and graphics.

1.

2. <http://www.oracle.com/technetwork/articles/datawarehouse/saternos-r-161569.html>

3. <http://www.r-project.org/>