



Revolution R Enterprise™ 7

Installation Guide for Teradata

The correct bibliographic citation for this manual is as follows: Revolution Analytics, Inc. 2014. *Revolution R Enterprise 7 Installation Guide for Teradata*. Revolution Analytics, Inc., Mountain View, CA.

Revolution R Enterprise 7 Installation Guide for Teradata

Copyright © 2014 Revolution Analytics, Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of Revolution Analytics.

U.S. Government Restricted Rights Notice: Use, duplication, or disclosure of this software and related documentation by the Government is subject to restrictions as set forth in subdivision (c) (1) (ii) of The Rights in Technical Data and Computer Software clause at 52.227-7013.

Revolution R, Revolution R Enterprise, RPE, RevoScaleR, DeployR, RevoTreeView, and Revolution Analytics are trademarks of Revolution Analytics.

Other product names mentioned herein are used for identification purposes only and may be trademarks of their respective owners.

Revolution Analytics
2570 West El Camino Real
Suite 222
Mountain View, CA 94040
U.S.A.

Revised on December 1, 2014

We want our documentation to be useful, and we want it to address your needs. If you have comments on this or any Revolution document, write to doc@revolutionanalytics.com.

Table of Contents

1	Quick Overview.....	5
2	System Requirements	6
3	Installing the Revolution R Enterprise rpms	7
4	Adding and Configuring Users	8
5	Managing Memory in In-Teradata Computations.....	9
6	Installing Additional R Packages	10
7	Removing Revolution R.....	11

1 Quick Overview

Revolution R Enterprise for Teradata is an R-based analytical engine embedded in your Teradata data warehouse. Together with a Revolution R Enterprise client, it provides a comprehensive set of tools for interacting with the Teradata database and performing in-database analytics. This manual provides detailed instructions for installing Revolution R Enterprise for Teradata.

NOTE: Revolution R Enterprise for Teradata is required for running Revolution R Enterprise scalable analytics in-database. If you do not need to run your analytics in-database, but simply need to access Teradata data via Teradata Parallel Transport or ODBC, you do not need to install Revolution R Enterprise in your Teradata data warehouse.

ATTENTION TERADATA EMPLOYEES AND CONTRACTORS. Do not proceed to download or install this software if you are an employee, contractor or other representative of Teradata Corporation or any of its affiliates, *except when installing this software on a system owned by, and for the internal use of, Teradata Corporation or any of its affiliates.*

Before installing Revolution R Enterprise for Teradata, you must have a fully configured Teradata data warehouse running either SUSE Linux Enterprise Server 10 (SLES 10) or SUSE Linux Enterprise Server 11 (SLES 11).

1. Download the Revolution R Enterprise distribution appropriate for your Linux distribution:

```
Revo-Ent-7.3.0-TERA14S10.tar.gz
```

```
Revo-Ent-7.3.0-TERA14S11.tar.gz
```

2. Unpack the distribution using the tar command as follows:

```
tar -zxf Revo-Ent-7.3.0-TERA14S10.tar.gz  
or  
tar -zxf Revo-Ent-7.3.0-TERA14S11.tar.gz
```

3. Change directory to the RevolutionR_7.3.0 directory created in Step 2:

```
cd RevolutionR_7.3.0
```

4. Run the install script, install.py, as follows:

```
./install.py
```

This script asks you to accept the Revolution R Enterprise license, and, if you are a Teradata employee, to acknowledge that, if you are installing the software, it is on a system owned by and for the internal use of Teradata Corporation or any of

its affiliates. It then copies the Revolution packages to the proper location for subsequent installation by the Teradata Parallel Upgrade Tool (PUT).

5. To install the Revolution R Enterprise rpms, use the Teradata Parallel Upgrade Tool (PUT). A quick guide to the required steps is provided in Chapter 3, Installing the Revolution R Enterprise rpms. For a more complete description, see the Teradata manual [Parallel Upgrade Tool \(PUT\) Reference](#).
6. After you have installed the rpms, change directory to the "teradata" directory beneath the RevolutionR_7.3.0 directory, and run the revoTdSetup.sh script as follows:

```
cd teradata
./revoTdSetup.sh
```

Enter your database administrator user name and password at the prompts. Do *not* modify this script, and in particular, do *not* modify the name of the revoAnalytics_Zqht2 database. The database administrator running the script must have specific grants, as described in Chapter 4, Adding and Configuring Users.

7. Create additional users and databases (for example, the RevoTestDB database to hold sample Revolution data) and grant them appropriate permissions as described in Chapter 4, Adding and Configuring Users. All grants (or equivalents) listed in Chapter 4 are required. Grants *must* be given to revoAnalytics_Zqht2 in order for RRE's in-database procedures and functions to read and write data, as required to carry out the user's in-database requests.
8. Stop and restart the Teradata database.

2 System Requirements

Revolution R Enterprise for Teradata has the following system requirements:

Processor: 64-bit processor with x86-compatible architecture (variously known as AMD64, Intel64, x86-64, IA-32e, EM64T, or x64 chips). Itanium-architecture chips (also known as IA-64) are not supported. Multiple-core chips are recommended.

Operating System: SUSE Linux Enterprise Server 10 or 11. Only 64-bit operating systems are supported.

Teradata Version: Teradata 14.10 (specifically, 14.10.01.07-1 or later in the .01 series, or 14.10.02d later in the .02 series).

Memory: A minimum of 1GB of RAM is required; 4GB or more are recommended.

Disk Space: A minimum of 500MB of disk space is required

3 Installing the Revolution R Enterprise rpms

In this chapter, we discuss the basics of using the Teradata Parallel Update Tool (PUT) to install the Revolution R Enterprise rpms.

PUT is a browser-based utility. To get started, point your Java-enabled browser the following URL:

<https://<HOSTIP>:8443/put>

where <HOSTIP> is the IP address of your Teradata data warehouse node and log in using a *Linux* account such as root (*not* a database account). To install the Revolution R Enterprise rpms on all the nodes, do the following:

1. The initial PUT screen asks you to select an operation. Click **Install/Upgrade Software**.
2. PUT starts you at the **Configuration Mode** screen. Verify that the Typical configuration mode is selected, and then click **Next**.
3. On the **Find All Subnets** screen, find the subnet matching your Teradata installation, and then click **Next**.
4. On the **Select Nodes** screen, select the nodes on which you want to install the Revolution R Enterprise rpms (typically the RevolutionR_7.3.0/rpms directory), and then click **Next**.
5. Continue accepting defaults and clicking Next as requested until you reach the **Enter Source for New Packages** screen. From this screen, navigate to the directory containing the Revolution R Enterprise rpms. Click **Next**.
6. On the **Media Source(s)** screen, verify that **No** is selected, and then click **Next**.
7. On the **Media Source(s) To Be Scanned** screen, click **Next**.
8. On the **Group Nodes** screen, ensure all the desired nodes are listed, and then click **Next**.
9. On the **Select Packages** screen, select **Revo_Packages_GPL**. Copy all nodes to the **Install To** box, and then click **Next**, and then, when asked if you want PUT to auto select these packages, click **Yes**.
10. On the **Select Packages** screen, select **Revo_Packages**. Again, copy all nodes to the **Install To** box, and then click **Next**.
11. On the **Select Packages** screen, select **R_GPL-devel**. Again, copy all nodes to the **Install To** box, and then click **Next**.
12. Click **Next**.
13. On the **Confirm Package Selection** screen, click **Next**.
14. Verify that the **Non-VM&F upgrade - use Primary BE** option is selected, and then click **Next**.
15. When the **Begin Install Packages Now?** Screen appears, click **Next**. You should see a progress screen telling you that your rpms are being installed on each of the nodes.
16. Continue clicking **Next**, accepting defaults, until you reach the screen containing the option **Automatically Restart the Database after Reboot**. Set this option according to your local requirements.

4 Adding and Configuring Users

All users who will interact with the database, including the database administrator, need to have specific grants. The following example creates a database administrator user named "sysdba" with the required grants to run the `revoTdSetup.sh` script:

```
# Create the user who will run revoTdSetup.sh (e.g., sysdba):
bteq .logon localhost/dbc,dbc
create user SYSDBA from dbc as perm=25E9, spool=15E9, temporary=15E9, password=sysdba;
grant select, execute on dbc to sysdba with grant option;
grant all on sysdba to sysdba with grant option;
grant select on sys_calendar to sysdba with grant option;
grant select on dbc to public;
```

The included script `revoTdUserSetup.bteq`, included in the `teradata` directory beneath the `RevolutionR_7.3.0` directory created when you unpack the installer tarball, outlines the basic steps for creating new database users and granting those user permissions to run in-database with their own individual databases. Copy and modify this script as necessary to work with your database configuration.

You will often want to allow a user account to run jobs on data in other databases also.

In addition, after re-creating the database as described below, you will need to re-grant database permissions to your user account in order to run jobs with it. Follow the steps below to do so. (For convenience, these steps are organized according to the purpose of the permissions granted.)

1. First, you will need to grant permissions for the user account to work with the Revolution database. Logon to `bteq` with your admin account, and run the following lines. Substitute the new user account name for 'ut1' in the sample lines below.

```
-- grant ut1 various permissions on revoAnalytics_Zqht2 DB
grant execute procedure on revoAnalytics_Zqht2.InitRevoJob to ut1;
grant execute procedure on revoAnalytics_Zqht2.StartRevoJobXsp to ut1;
grant execute procedure on revoAnalytics_Zqht2.GetRevoJobState to ut1;
grant execute procedure on revoAnalytics_Zqht2.RemoveRevoJobResults to ut1;
grant execute function on revoAnalytics_Zqht2.FinishRevoJob to ut1;
grant execute function on revoAnalytics_Zqht2.DistributeRevoJobDataUdf to ut1;
grant execute function on revoAnalytics_Zqht2.revoAnalyticsTblOp to ut1;
grant select on revoAnalytics_Zqht2.RevoAnalyticsJobResults to ut1;

-- the following two create view and drop view grants are for the
-- table operator PARTITION BY work-around
-- for Teradata's Issue DR166609, which was fixed in 14.10.02d
grant create view on revoAnalytics_Zqht2 to ut1;
grant drop view on revoAnalytics_Zqht2 to ut1;
```

2. Next you will need to grant permissions for the user account and the revolution database to work with data in the user account database. Run the following lines, substituting your user account name for 'ut1'.

```
-- grant ut1 rights on db with data to be analyzed -
-- in this case itself
grant all on ut1 to ut1;
```

```
-- give revoAnalytics_Zqht2 rights on database with data to
-- be analyzed - in this case ut1
grant select on ut1 to revoAnalytics_Zqht2;
grant create table on ut1 to revoAnalytics_Zqht2;
```

3. Finally, you will want to grant permissions on any other databases you wish that account to have access to. Run the following lines, substituting your user account name for 'ut1', and your other database name for 'RevoTestDB'.

```
-- grant ut1 rights on db with data to be analyzed
grant all on RevoTestDB to ut1;

-- give revoAnalytics_Zqht2 rights on database with data to
-- be analyzed
grant select on RevoTestDB to revoAnalytics_Zqht2;
grant create table on RevoTestDB to revoAnalytics_Zqht2;

-- this last grant is required if you want to be able to
-- overwrite output tables created using TPT with a local
-- context - note that some administrators may not
-- be comfortable giving our software this grant as it means
-- a bug could destroy existing data
grant drop table on RevoTestDB to revoAnalytics_Zqht2;
```

5 Managing Memory in In-Teradata Computations

Because a Teradata cluster node typically has many worker processes (AMPs) running constantly, it is important to limit how much memory a distributed analysis consumes.

Revolution R Enterprise 7.3.0 and later includes both default and configurable memory limits for master and worker processes working on a RevoScaleR job. By default, the master process memory limit is 2000000000 bytes (approximately 2GB) and the worker process memory limit is 1000000000 (approximately 1GB).

This may be customized by the cluster administrator using the environment variables REVOMASTERMEMLIMIT and REVOWORKERMEMLIMIT. These environment variables can be defined in many places, but are probably best defined by creating (or editing if already created) the Renviron.site file on each node of the cluster. Setting them in Renviron.site limits them to the processes that will be using them, which reduces the chances of variable name collision.

Important: Do not confuse the Renviron.site file with the Renviron file (no extension) that is found in the same directory (that is, the R 'etc' directory). The Renviron file is read very early in the R startup process and is not intended to be edited by anyone but a member of R core. The Renviron.site file does not normally exist until someone at your site creates it, but is the recommended location for system-wide changes to your R environment variable settings.

10 Installing Additional R Packages

You can create or locate `Renviron.site` in the R 'etc' directory on the node, typically `/usr/lib64/Revo-7.3/R-3.1.1/lib64/R/etc/`. For example:

```
vi /usr/lib64/Revo-7.3/R-3.1.1/lib64/R/etc/Renviron.site
```

Add two lines to this file:

```
REVOMASTERMEMLIMIT=?  
REVOWORKERMEMLIMIT=?
```

where ? is the limit you wish to set on these processes in bytes.

Save the file, then ensure that the Teradata `tdatuser` account has read access to the file. You can do this with `chmod` as follows:

```
chmod a+r /usr/lib64/Revo-7.3/R-3.1.1/lib64/R/etc/Renviron.site
```

After creating or changing these environment variables, the database administrator must restart the Teradata database before the new limits will take effect.

6 Installing Additional R Packages

The R community is an active, open-source community, and new packages extending R's capacity for statistics, data analysis, graphics, and interconnectivity are added frequently. The most up-to-date source for these third-party packages is the Comprehensive R Archive Network (CRAN), a network of servers around the world that store open-source R distributions, extensions, documentation and binaries. The repository (<http://cran.r-project.org/>) has grown from only 12 packages in 1997 to over 5300 packages currently. However, CRAN does not maintain source packages by R version, and Revolution R Enterprise for Teradata is seldom the latest R version, so packages from CRAN may be incompatible with Revolution R Enterprise for Teradata. Revolution Analytics does maintain source packages by version on its own repository, <http://packages.revolutionanalytics.com>. Users may take advantage of this repository and download the chosen packages, but notice that they are completely external to Teradata.

In most cases, the natural place to install additional R packages is to the client workstation. Installation to the Teradata data warehouse is required only if you plan to use the package's functions as transform functions inside `RevoScaleR` functions, in which case the packages will actually need to be loaded in-database. If you need a package for this purpose, you can install it in-database in either of two ways:

1. Use PUT together with the included `Rpkg2rpm` script
2. Manually distribute and install the package on the nodes of your data warehouse.

To use PUT together with the included `Rpkg2rpm` script:

1. Download the package to your Revolution installation directory. For example, to install the SuppDists package of supplemental statistical distributions, you would get the package as follows:

```
wget packages.revolutionanalytics.com/cran/3.1/stable/src/contrib/SuppDists_1.1-9.1.tar.gz
```

2. Run the Rpkg2rpm.py script as follows:

```
python Rpkg2rpm.py SuppDists_1.1-9.1.tar.gz
```

3. Then run the PUT utility again to distribute and install the package.

To manually distribute and install the package:

1. Download the package and any required dependencies from <http://packages.revolutionanalytics.com>.
2. Copy the downloaded packages to each node of your data warehouse.
3. For each package, run the command “R CMD INSTALL *package.tar.gz*” on each node. (If your data warehouse is equipped with the psh command, you can use that to run the command on all the nodes in parallel.)

7 Removing Revolution R

To remove Revolution R Enterprise from your computer, run the following commands:

```
psh zypper --verbose --non-interactive remove Revo_Packages_GPL
psh zypper --verbose --non-interactive remove Revo_Packages
psh zypper --verbose --non-interactive remove R_GPL-devel
psh zypper --verbose --non-interactive remove R_GPL
psh zypper --verbose --non-interactive remove Revo_Core
```