Installation of Open Source R
on the
IBM PureData for Analytics Appliance

The intent of this document is to enable the use of open source R as an in-database executable language for the IBM PureData for Analytics (PDA) appliance. This document describes how you can download, compile and install Open Source R and some prerequisite tools on the PDA appliance by using the *installOpenSourceR.sh* shell script.

It is recommended to first install Open Source R on a PDA emulator (the directions and code are designed for this approach). The compiled code can be easily transferred to other compatible appliances or emulators, so typically, it does not matter where the source code compilation is done. You can download an emulator from the IBM developerWorks Netezza Developer Network (NDN) community. You need to register first at developerWorks (https://www.ibm.com/developerWorks). Search for “NDN” to locate the Netezza Developer Network community. On the “Community Description” page you will find a link for “Development Tools”. Selecting this will take you to a page with information on the PDA emulator as well as a link to download for the emulator.

NOTE: In some environments, IT regulations do not allow emulators. In such situations, one can build R in the physical Netezza appliance directly, however the emulator is still the recommended environment for this task because of the required root access to the SPU/Blades.

Before you can use the installOpenSourceR.sh script, you must install IBM Netezza Analytics (INZA) since the installation of Open Source R requires several tools that are installed with INZA. For each version of INZA, a corresponding R version is recommended as follows. Other combinations might be possible, but have not been tested.

- INZA v2.5.0 to 2.5.3: R 2.14.x
- INZA v2.5.4 to 3.0.0: R 2.15.x
- INZA v3.0.1 to 3.0.2: R 3.0.x
- INZA v3.2.x or greater: R 3.x

Installation Overview

The installation is a straightforward process when your emulator has internet access:

1. Copy the installation script to the emulator
2. Download the desired version of R and copy to the emulator

3. Run the installation script to download dependencies and compile/build open source R for PDA

4. Copy the newly created R binaries to other systems needing the capability to run open source R in-database.

If the emulator does not have internet access, step three will be preceded with a step to manually download several software dependencies and copy them to the emulator before the installation script is executed.

Installation Requirements

- Installing more recent versions of R
    - The installation script will automatically install R v3.0.2. If you desire a different, more recent version, you must download the source code for that version and place it in the correct destination directory in the emulator. If a version greater than R v3.0.2 is found by the installation scripts, this version will be installed.
    - This functionality requires the use of INZA 3.2.x or greater.
- Internet Access
    - They emulator being used to build open source R for PDA requires internet access. The installation script will automatically download linux software dependencies for R if they are not installed on the system.
    - If providing internet access from the emulator is not possible, the software dependencies can be downloaded manually and transferred to the system. Appendix 1 provides details on these dependencies.
    - It is highly recommended that an emulator with internet access is used for the compile/build/installation procedure.
- Administrative system access
    - You need the root password for the SPUs because some code must be compiled on a SPU. The root password for the host is not needed. A*ll operations on the host are performed by user nz, including the execution of installOpenSourceR.sh*.
    - It is highly recommended that the compile/build process for R be performed on an emulator. The resultant binaries can then be installed on other systems.
    - It is recommended that a new version of R be built for each NPS/INZA version combination that is being used.
    - Administrator access is required for the final installation on target systems after the initial build on the emulator.

Build and Install Open Source R for PDA on the Emulator

In the following directions, the letter “x” is used in various file and folder names. When used in the R download from CRAN (*R-3.x.x.tar)*, it is a variable for the specific version you are downloading. When used in the file name for the *Netezza_R_Installation_and_Client_R-3.vx.zip* and associated directory name, it is referring to the version of the Open Source R Installation from NDN.

1) Copy the *Netezza_R_Installation_and_Client_R-3.vx.zip* file to the */nz/var/inza* directory.
2) Unzip the *Netezza_R_Installation_and_Client_R-3_vx.zip* file:
   - gunzip Netezza_R_Installation_and_Client_R-3_vx.zip
3) Download the desired open source R version from the CRAN repository (https://cran.r-project.org/src/base/R-3/) and copy the open source R download (*R-3.x.x.tar.gz*) to the */nz/var/inza/Netezza_R_Installation_and_Client_R-3_vx* directory.
4) Move to the newly created directory
   - cd /nz/var/inza/Netezza_R_Installation_and_Client_R-3_vx
5) To see the available options, you can first run the script with the --help option:
   - ./installOpenSourceR.sh --help
6) Typically, the only option you need is the option to specify the SPU root password:
   - ./installOpenSourceR.sh -p <spu-root-password>

Depending on the performance of your emulator the whole compilation and installation process can take more than an hour.

Uninstall Open Source R for PDA

To uninstall Open Source R, call the script by entering the following command:

- ./installOpenSourceR.sh -u

Use Binaries to Install Open Source R for PDA on Target Machines

After the binaries are compiled and installed on one machine, you can install the compiled binaries on another PDA system, hereafter called *target machine*.

Prerequisite: Netezza Analytics (INZA) must be installed on the target machine. INZA can be downloaded from IBM Fix Central.

To install the binaries on the target machine, do the following steps:

1. Create the following directory on the target machine. The directory name must be identical to the directory name on the source machine.
   - /nz/extensions/nz/r_install
2. From the /nz/extensions/nz/r_install directory on the source machine, copy the following files to the newly created directory on the target machine:
   - sysroot_overlay.tar.gz
   - r_<version>_overlay.tar.gz

3. Copy also this script *installOpenSourceR.sh* to this directory.
4. Call the *installOpenSourceR.sh* script by entering the following command. The SPU password is not required in this case:
    - ./installOpenSourceR.sh -i -b

Uninstall Open Source R for PDA from Target Machines

To uninstall Open Source R, call the script by entering the following command:

- ./installOpenSourceR.sh -u

Testing the Open Source R for PDA Installation

An extensive test of the R installation should be done from an R client that has been extended with the Netezza Analytics R client packages (nzr, nza, nzmatrix). The *demo* functions of these packages also call R functions on the server. The installation and usage of the Netezza Analytics R client packages is described in the client package documentation.

However, for a quick test, you can run the following SQL command from the nzsql prompt on the emulator and/or target system after installation. This will validate that the in-database version of R is running properly:

*SELECT * FROM TABLE WITH FINAL(nzr..r_udtf('CODE_PLAIN="getNext();setOutput(0,"output value");outputResult()'));*

Appendix I – Manual Installation of Software Dependencies

Several Linux tools must be installed before you can install Open Source R. All these tools are installed by the script automatically. You just need to download all these source code archives from the corresponding websites. The following table lists the tools and recommended URLs. Store all the downloaded files and the *installOpenSourceR.sh* script in the same directory on the Netezza host.

If the script cannot find one of the tools or OpenSource R, it tries to download the source code itself, using the URLs from the table below. For bash, it also downloads all patches it can find, to avoid bash security problems that were found for the “shellshock” vulnerability.

Source Code Example URL

ftp://ftp.gnu.org/pub/gnu/termcap/termcap-1.3.1.tar.gz
ftp://ftp.gnu.org/pub/gnu/ncurses/ncurses-5.9.tar.gz
ftp://ftp.cwru.edu/pub/bash/readline-6.2.tar.gz
ftp://ftp.gnu.org/pub/gnu/libiconv/libiconv-1.14.tar.gz
ftp://ftp.gnu.org/pub/gnu/make/make-3.82.tar.gz
ftp://ftp.gnu.org/gnu/bash/bash-4.3.tar.gz
ftp://ftp.gnu.org/gnu/coreutils/coreutils-8.13.tar.gz
ftp://ftp.gnu.org/gnu/findutils/findutils-4.4.2.tar.gz
ftp://ftp.gnu.org/gnu/grep/grep-2.9.tar.gz
ftp://ftp.gnu.org/gnu/sed/sed-4.2.tar.gz
ftp://ftp.gnu.org/gnu/tar/tar-1.26.tar.gz
http://xcb.freedesktop.org/dist/libpthread-stubs-0.1.tar.bz2
http://pkgconfig.freedesktop.org/releases/pkg-config-0.28.tar.gz
http://zlib.net/zlib-1.2.8.tar.gz
https://sourceforge.net/projects/pcre/files/pcre/8.38/pcre-8.38.tar.gz/download
http://www.bzip.org/1.0.6/bzip2-1.0.6.tar.gz
http://tukaani.org/xz/xz-5.2.2.tar.bz2
https://www.openssl.org/source/openssl-1.0.2h.tar.gz
https://curl.haxx.se/download/curl-7.49.1.tar.bz2

http://lib.stat.cmu.edu/R/CRAN/src/base/R-2/R-2.14.2.tar.gz
or
http://lib.stat.cmu.edu/R/CRAN/src/base/R-2/R-2.15.3.tar.gz
or
http://lib.stat.cmu.edu/R/CRAN/src/base/R-3/R-3.0.2.tar.gz
or
http://lib.stat.cmu.edu/R/CRAN/src/base/R-3/R-3.1.x.tar.gz
or
http://lib.stat.cmu.edu/R/CRAN/src/base/R-3/R-3.2.x.tar.gz
or
http://lib.stat.cmu.edu/R/CRAN/src/base/R-3/R-3.3.x.tar.gz

Copyright and trademark information