CARMAweb installation instructions 1.0.0

Johannes Rainer

March 2, 2006

Contents

List of Figures

Installation instructions

The archive `carmaweb.tar.gz` contains the following files:

- `R-Install`: contains required packages and scripts to install the packages needed for CARMAweb (see section 0.2).

- `bin`: contains simple bash scripts to start Rserve and CARMAweb.

- `jboss-4.0.2`: the application server that runs CARMAweb.

- `data`: sub directories where user data will be stored into (CARMAweb can also be configured to store data in other places.)

- `users.xml`: file that will be used for the usermanagement to store user information into.

In the installation instructions below we are assuming that the archive is unpacked into the home directory of the user that will be used later to run CARMAweb (e.g. the archive has been copied to the users home directory (e.g. */home/carma*) and is unpacked using the command tar -xzf carmaweb.tar.gz).

0.1 Requirements

Unfortunately Microsoft Windows is not supported yet as OS because some BioConductor packages are not available for Windows.

0.1.1 graphviz

`graphviz` is available for most Linux distributions as an rpm package (note that also the development package is needed!), otherwise it can be downloaded from http://www.graphviz.org/. The graphviz software is needed for the `Rgraphviz` BioConductor package that is used to draw the GO (GeneOntology) graphs.

0.1.2   postgresql

The postgresql libraries and headers have to be installed on the system where CARMAweb should be installed onto. For most linux distributions these are provided by rpm packages.

0.2   Installing R and Bioconductor

CARMAweb should in principle be able to use all versions of $R >= 2.2$. R can be downloaded from http://cran.r-project.org/. If `R` is compiled from source the configure flag `--enable-R-shlib` has to be used, otherwise `Rserve` can not be installed.

0.2.1   Rserve

Rserve is used by CARMAweb to communicate with R. The Rserve package (`Rserve_0.3-17.tar.gz`) can be found in the directory `carmaweb/R-Install` or can be downloaded from http://stats.math.uni-augsburg.de/Rserve/ . The easiest way to install this package is to change to the directory where it is located and type as superuser (!!!) the command

```
R CMD INSTALL Rserve_0.3-17.tar.gz
```

Usually there should be no problems installing this package. Specific Rserve installation instructions can also be found on that page.

All further instruction in this section assume that you are logged in as the user that runs CARMAweb (e.g. if CARMAweb should run as user *carma* log in as user *carma*. It is not recommended to run CARMAweb as superuser!). For the *not-system-wide* installation of R packages copy the `.Renviron` file from the directory `R-Install` into the home directory of the user that will run CARMAweb. By default this will cause all additional packages to be installed into the directory `<USER_HOME>/R/library`. Copy also the file `.Rprofile` into the home directory of the user that will run CARMAweb. Additionally some system variables have to be defined by editing the `.bash_profile` file (or any other profile file where system variables can be declared) of the user (located in his user directory). Add the `R_LIBS` variable to point to the local users library installation and to the system R libraries (e.g. `R_LIBS=~/R/library:/usr/lib/R/library`). It is also recommended to set the environment variable `LD_LIBRARY_PATH` to point also to the `graphviz` installation. For example add the line

```
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/usr/lib/graphviz
```

(if graphviz was installed in `/usr/lib/graphviz`) to the `.bash_profile` file in the users home directory (remember also to export this variable by adding the line `export LD_LIBRARY_PATH` in the file and to reload the profile (`source ~/.bash_profile`)).

By adding the following lines to the `.bash_profile` file (or any other file where system variables can be declared), we are telling R that the user has additional packages installed in the directory *R/library* in his home

directory and that the core libraries from R are located at */usr/lib/R/library*.

```
R_LIBS=~/R/library:/usr/lib/R/library
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/usr/lib/graphviz


export R_LIBS LD_LIBRARY_PATH
```

0.2.2   Installing required Bioconductor packages

The current version of CARMAweb (1.0.0) has been tested successfully with Bioconductor version 1.7, but also later versions should work. (Note: it is assumed that you are logged in as the user that will be used to run CARMAweb!).

- Type `echo $R_LIBS` in the console, if it does not return something like *~/R/library* the system variables are not set and the profile file has to be reloaded (e.g. by typing `source ~/.bash_profile`).

- Change into the `R-Install` directory and type `R CMD INSTALL reposTools_1.8.0.tar.gz`, this will install the Bioconductor package `reposTools` that will be used to install all required packages.

- Next the package *xlahomology* has to be installed. This can be done by typing
  `R CMD INSTALL xlahomology_1.8.4.tar.gz`.

- Next copy all base packages from the R installation to `~/R/library` (e.g. all directories from the `/usr/lib/R/library` directory (or `/usr/lib64/R/library` to `~/R/library`)). This can be performed for example with the command
  `cp -R /usr/lib/R/library/ ~/R/`

- Change again into the `R-Install` directory and start `R` (by typing *R* in the console)

- Type
  `source("install.R")`
  in the R console

- Type
  `install()`
  This command installs all needed Bioconductor packages and all annotation and metadata packages. The needed Bioconductor packages are specified in the file *Rweb_packages*, whereas the annotation packages are listed in the *chips.txt* file. As this function installs the required packages including all packages the required packages depend on it takes several minutes (or also hours, depending on the speed of the internet connection) to finish.

- At last quit R (*q()*) and finally install the latest version of the `maDB` package by typing

  `R CMD INSTALL maDB_1.3.7.tar.gz`.

It is important to note that there should be no errors during the installation of required Bioconductor packages! It is not suggested to go on with the installation of CARMAweb if not all required packages where installed correctly!

0.3 Java

CARMAweb requires Java Development Kit (JDK) >= 1.4.2. The environment variable `JAVA_HOME` has to point to the JDK installation directory, `JAVA_HOME/bin` has to be in the `PATH` environment variable.

0.4 Installation of CARMAweb

The web application CARMAweb and all other components (e.g. the cluster analysis module GenesisWeb) run within the application server jboss (http://www.jboss.org). CARMAweb is highly configurable, but this also means that a lot of settings and variables have to be adjusted for CARMAweb to run without errors.

The *bin* directory in the directory *carmaweb* contains two shell scripts to start the application server (and therefore the web application) and the `Rserve` service.

The file `jboss_init.sh` has to be edited to be able to start CARMAweb. The `JBOSSDIR` attribute in the `jboss_init.sh` file has to be set correctly to point to the `bin` directory of jboss (e.g. to */home/carma/carmaweb/jboss-4.0.2/bin* if the carmaweb archive has been extracted to the home directory of a user called *carma*). Another attribute that should be adjusted is the `LOGFILE` attribute that specifies where the application should write log messages into.

The `startRserve.sh` is used to start the `Rserve` service. If a system variable called `R_HOME` is already correctly defined nothing has to be configured in this file, otherwise the `R_HOME` attribute has to be set correctly in this file (to point to the R installation directory (usually one of */usr/lib/R*, */usr/lib64/R*, */usr/local/lib/R* or */usr/local/lib64/R*)).

0.4.1 What ports are used?

The default ports that are already set for CARMAweb in the jboss application server are:

- 38080 to answer HTTP requests

- 31099 for the JNDI port

If you are using a firewall only the 38080 port has to be opened, as it is used to answer all incoming requests.

0.4.2   Configuration of CARMAweb

In this section we will go through all CARMAweb modules and see how and where these modules have to be configured. All modules are located within the directory `jboss-4.0.2/server/Rweb/deploy`, therefore all files described in this section are relative to this path.

0.4.2.1   CheckRServe.sar

This module controls that the `Rserve` is running properly and that it accepts connections. Therefore it tests periodically if `Rserve` is up and running. The configuration file for this service is *CheckRServe.sar/META-INF/jboss-service.xml* (located in the *deploy* directory of the jboss application server). This file has to be configured by specifying where the shell script file to start `Rserve` is located (edit the *StartCall* attribute to point to the correct location (by default it assumes that the shell script to start `Rserve` is located at */home/carma/carmaweb/bin/startRserve.sh*)).

0.4.2.2   ClusterService.sar

This service allows to perform different ways of load balancing by assigning the various jobs to different computers in a cluster. The configuration file for this service is located at *ClusterService.sar/META-INF/jboss-service.xml*. For simple (not load balancing purpose) the only parameter that has to be set correctly is the *PathToSavedFiles* attribute. Set this correctly to a directory where temporary cluster analysis results should be stored to (defaults to */home/carma/carmaweb/data/genesisweb/SavedJOBS*).

0.4.2.3   RServer.ear

This module is used to establish asynchroneous connections between the CARMAweb web application and the `Rserve` service. The configuration file to configure this service is *RServer.ear/RServer.jar/RServe.properties*. The parameters that have to be adjusted are:

- `workspace.path`: used to specify the directory where the workspaces for SOAP calls should be stored to.

- `R.path`: the directory where R has been installed to (usually */usr/lib/R* or */usr/lib64/R*).

- `maxtime`: SOAP workspaces older than the time specified (in milliseconds) will be deleted automatically. Defaults to 1 week (*maxtime=302400000*).

0.4.2.4   carma.war

The CARMAweb web application. This module has to be configured via the *carma.war/WEB-INF/web.xml* file. The parameters that have to be adjusted in this file are:

- CONFIG_FILE_PATH: has to point to the R installation directory (e.g. */usr/lib/R*)

- PATH_TO_R: same as *CONFIG_FILE_PATH*

- USER_DIRECTORY: the directory where all user data should be stored into (workspaces, data...). Defaults to */home/carma/carmaweb/data/carma*.

To enable the SOAP interface to allow also other programs to use features of CARMAweb the *carma.war/WEB-INF/server-config.wsdd* file has to be edited. The attachments.Directory has to be adjusted if carmaweb was not unpacked to */home/carma*. Actually just the first part (*/home/carma/*) of the default value has to be replaced by the directory where the carmaweb archive has been unpacked. If this parameter is not configured correctly, SOAP clients can not upload files to the CARMAweb web application.

0.4.2.5   genesis.war

The GenesisWeb web application. Again the *web.xml* file located in the directory *genesis.war/WEB-INF* has to be adjusted to configure this module. The only parameter to edit is the PATH_TO_USER_DIR attribute that has to point to the directory where the genesisweb user data should be stored into.

0.4.2.6   Configuring the usermanagement

CARMAweb comes with its own simple usermanagement.To configure it the files *bugblatterbeast-usermanagement-utils.properties* in the directories *carma.war/WEB-INF/classes* and *genesis.war/WEB-INF/classes* have to be re-named into $< HOSTNAME >$-*usermanagement-utils.properties*, where $< HOSTNAME >$ is the hostname of the server where CARMAweb has been installed into (just type *hostname* in a console to get the machines hostname). Additionally in both files the parameter sum_users_filepath has to be adjusted to point to the *users.xml* file that is usually located in the *carmaweb* directory.

0.5   Starting CARMAweb

To start CARMAweb (again assumed you are logged in as the correct user) change to the directory *carmaweb/bin* and type

```
./jboss_init.sh start
```

This should start the application server (and therefore also CARMAweb). Have a look at the log file (can be specified in 0.4). If there are no ERROR messages everything started successfully. It is not needed to start also Rserve manually, because it is started automatically during the CARMAweb startup. The startup of CARMAweb depends on the machine where it is installed, but usually it takes about 1 to 2 minutes.

To see if CARMAweb runs properly open a web browser and go to the url *http://localhost:38080/carma* and create your first account using the *Create account* link.