



CLC **Bioinformatics Database**

Administrator

USER MANUAL

Administrator Manual for
CLC Bioinformatics Database 5.0
Windows, macOS and Linux

November 28, 2018

This software is for research purposes only.

QIAGEN Aarhus
Silkeborgvej 2
Prismet
DK-8000 Aarhus C
Denmark



Contents

1	Introduction	6
1.1	System requirements	8
1.2	Licensing	8
1.3	Latest improvements	8
2	Installation	9
2.1	Quick installation guide	9
2.2	Installing the database	10
2.2.1	Download and install a Database Management System	10
2.2.2	Create a new database and user/role	10
2.2.3	Initialize the database	10
2.3	Installing and running the Server	11
2.3.1	Installing the Server software	12
2.4	Installation modes - console and silent	14
2.5	Upgrading an existing installation	14
2.5.1	Upgrading between major versions	15
2.6	Allowing access through your firewall	15
2.7	Downloading a license	15
2.7.1	Windows license download	15
2.7.2	macOS license download	16
2.7.3	Linux license download	16
2.7.4	Download a static license on a non-networked machine	16
2.8	Starting and stopping the server	17
2.8.1	Microsoft Windows	17
2.8.2	macOS	18

2.8.3 Linux	18
3 Configuration	20
3.1 Logging into the administrative interface	20
3.2 Adding locations for saving data	20
3.2.1 File system locations	20
Important points about CLC Server file system locations	22
3.2.2 Enabling and disabling internal compression of CLC data	22
3.2.3 Database locations	23
3.2.4 Rebuilding the index	24
3.3 Changing the listening port	24
3.4 Setting the amount of memory available for the JVM	24
3.5 HTTP settings	25
3.6 Deployment of server information to CLC Workbenches	25
4 Managing users and groups	26
4.1 Logging in the first time and changing the root password	26
4.2 User authentication using the web interface	27
4.2.1 Authentication options	27
Built-in authentication	27
LDAP directory	27
Active Directory	29
4.2.2 Managing users and groups using built-in authentication	29
4.2.3 User authentication using the Workbench	30
4.2.4 Managing users through the Workbench	30
4.2.5 Managing groups through the Workbench	31
5 Access privileges and permissions	33
5.1 Controlling access to CLC Server data	33
5.1.1 Setting permissions on a folder	34
5.1.2 Recycle bin	35
5.1.3 Technical notes about permissions and security	36
6 Queue	37

7	Status and management	38
7.1	User statistics	38
7.2	System statistics	39
7.3	Server maintenance	39
8	Audit log	41
8.1	Customized attributes on data locations	42
8.1.1	Filling in values	44
8.1.2	What happens when a clc object is copied to another data location?	45
8.1.3	Searching	46
9	Appendix	47
9.1	Troubleshooting	47
9.1.1	Check setup	47
9.1.2	Bug reporting	48
9.2	Database configurations	49
9.2.1	Installing JDBC drivers	49
9.2.2	Configuring MySQL	51
9.3	SSL and encryption	51
9.3.1	Enabling SSL on the server	52
	Creating a PKCS12 keystore file	52
9.3.2	Logging in using SSL from the Workbench	52
9.3.3	Logging in using SSL from the <i>CLC Server Command Line Tools</i>	53
9.4	Third party libraries	54
	Bibliography	55

Chapter 1

Introduction

Welcome to *CLC Bioinformatics Database 5.0*, a central element of the CLC product line enterprise solutions.

The latest version of the user manual can also be found in pdf format at <http://www.qiagenbioinformatics.com/support/manuals/>.

You can get an overview of the server solution in figure 1.1. The software depicted here is for research purposes only.

The technical set-up of *CLC Bioinformatics Database* involves a database back-end and a server middle layer that the Workbench connects to. This server layer also has a web interface, which is used for administration and which also has very basic capabilities for uploading and downloading non-NGS data, and basic functionality for browsing and searching for data. The primary way to work with data stored in the *CLC Bioinformatics Database* would be to use a CLC Workbench.



Figure 1.1: An overview of the server solution. Note that not all features are included with all license models.

1.1 System requirements

The system requirements of *CLC Bioinformatics Database* are:

Server operating system

- Windows 7, Windows 8, Windows 10, Windows Server 2012, and Windows Server 2016
- OS X 10.10, 10.11 and macOS 10.12, 10.13, 10.14
- Linux: RHEL 6.7 and later, SUSE Linux Enterprise Server 11 and later. The software is expected to run without problem on other recent Linux systems, but we do not guarantee this.
- CLC Servers in a given setup (e.g. a master CLC Server and its job nodes or grid nodes), must run on the same type of operating system.

Server hardware requirements

- Intel or AMD CPU required
- Computer power: 2 cores required. 8 cores recommended.
- Memory: 4 GB RAM required. 16 GB RAM recommended.
- Disk space: The amount of space required will depend on the types and volume of the data being analyzed.

1.2 Licensing

Two kinds of license can be involved in running analyses on the *CLC Bioinformatics Database*.

- **A license for the server software itself.** This is needed for running analyses via the server. The license will allow a certain number of open sessions. This refers to the number of active, individual log-ins from server clients such as Workbenches, the Command Line Tools, or the web interface to the server. The number of sessions is part of the agreement with QIAGEN when you purchase a license. The manual chapter about installation provides information about how to obtain and deploy the license for the server.
- **A license for the Workbench software.** A Workbench is used to connect to the server to view the data stored there. Find the user manuals and deployment manual for the Workbenches at <http://www.qiagenbioinformatics.com/support/manuals/>.

1.3 Latest improvements

A list of new features, improvements, bugfixes, and changes for the current version of *CLC Bioinformatics Database* can be found at:

<http://www.qiagenbioinformatics.com/products/clc-bioinformatics-database/latest-improvements/current-line/>

Chapter 2

Installation

2.1 Quick installation guide

The following describes briefly the steps needed to set up a *CLC Bioinformatics Database 5.0* with pointers to more detailed explanation of each step.

1. Download and run the server software installer file. When prompted during the installation process, choose to start the server (section 2.3).
2. Run the license download script distributed with the server software. This script can be found in the installation area of the software. (section 2.7). The script will automatically download a license file and place it in the server installation directory under the folder called `licenses`.
3. Restart the server (section 2.8).
4. Ensure the necessary port is open for access by client software for the server. The default port is 7777 .
5. Log into the server web administrative interface using a web browser using the username **root** and password **default** (section 3).
6. Change the root password (section 4.1).
7. Configure the authentication mechanism and optionally set up users and groups (section 4.2).
8. Add data locations (section 3.2).
9. Check your server setup using the **Check set-up** link in the upper right corner as described in section 9.1.1.
10. Your server should now be ready for client software to connect to and use.

2.2 Installing the database

2.2.1 Download and install a Database Management System

If you do not have access to an existing installation of a Database Management System (DBMS) you will have to download and install one. *CLC Bioinformatics Database* can be used with a number of different DBMS implementations. Choosing the right one for you and your organization depends on many factors such as price, performance, scalability, security, platform-support, etc.

Information about the supported solutions are available via the links below.

- MySQL: <http://dev.mysql.com/downloads/>
- PostgreSQL: <http://www.postgresql.org/>
- Microsoft SQL Server: <http://www.microsoft.com/SQL/>
- Oracle: <http://www.oracle.com/>

You will need to make the appropriate JDBC driver available to the *CLC Bioinformatics Database*. See section 9.2 for details and for additional configuration information for certain DBMSs.

2.2.2 Create a new database and user/role

Once your DBMS is installed and running you will need to create a database for containing your CLC data. We also recommend that you create a special database-user (sometimes called a database-role) for accessing this database.

Consult the documentation of your DBMS for information about creating databases and managing users/roles.

2.2.3 Initialize the database

Before you can connect to your database from a CLC Workbench or Server it must be initialized. The initialization creates the required tables for holding objects, and prepares an index used for searching. Initialization is performed with the CLC Bioinformatics Database Tool (see figure 2.1).

- Download the CLC Bioinformatics Database Tool from <https://www.qiagenbioinformatics.com/products/clc-bioinformatics-database-tool-direct-download/>
- Install the CLC Bioinformatics Database Tool on a client machine, and start the program.
- Fill in the fields with the required information.
 - Hostname: The fully-qualified hostname of the server running the database.
NOTE: The same hostname must be used every time you connect to the database
 - Port: The TCP/IP listening port on the database server
 - Database name: The name of the database you created in the previous section
 - Username: the name of the user/role you created in the previous section

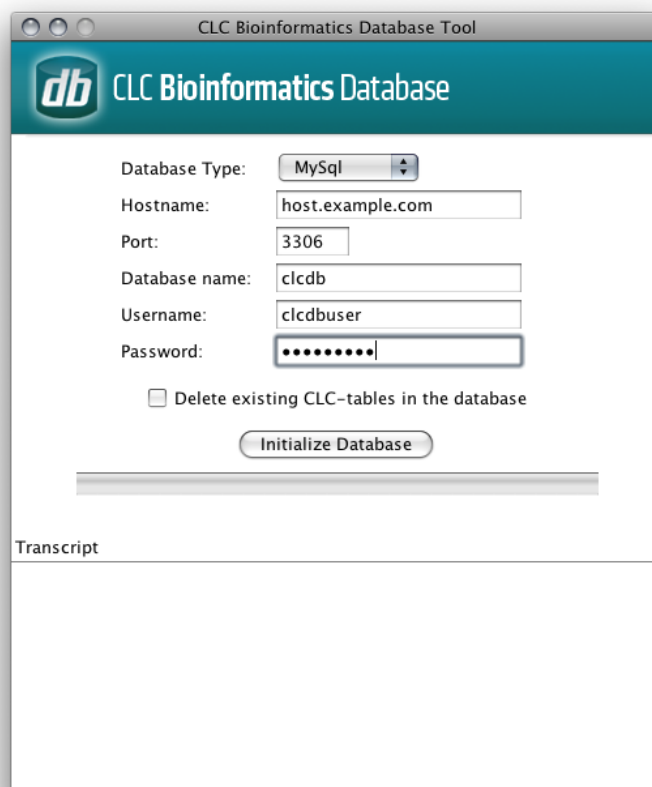


Figure 2.1: The CLC Bioinformatics Database tool

- Password: the password for the user/role.
- To re-initializing an existing CLC database you must check the "Delete Existing..." checkbox.
NOTE: ANY DATA ALREADY STORED IN THE CLC DATABASE WILL BE DELETED.
- Click the Initialize Database button to start the process.

While the program is working the progress-bar will show the status and the transcript will show a log of actions, events and problems. If anything goes wrong, please consult the transcript for more information. If you need assistance, please contact ts-bioinformatics@qiagen.com, and include the contents of transcript.

If the initialization is successful, the status bar will display this message: *Database successfully initialized*. You can now close the CLC Bioinformatics Database Tool.

2.3 Installing and running the Server

Getting the *CLC Bioinformatics Database* software installed and running involves, at minimum, these steps:

1. Install the software.
2. Ensure the necessary port in the firewall is open.

3. Download a license.
4. Start the Server and/or configure it as a service.

All these steps are covered in this section of the manual.

Installing and running the *CLC Bioinformatics Database* is straightforward. However, if you do run into troubles, please refer to the troubleshooting section in Appendix 9.1, which provides tips on how to troubleshoot problems yourself, as well as how to get help.

2.3.1 Installing the Server software

The installation can only be performed by a user with administrative privileges. On some operating systems, you can double click on the installer file icon to begin installation. Depending on your operating system you may be prompted for your password (as shown in figure 2.2) or asked to allow the installation to be performed.

- On Windows 8 and Windows 7, you will need to right click on the installer file icon, and choose to **Run as administrator**.
- For the Linux-based installation script, you would normally wish to install to a central location, which will involve running the installation script as an administrative user - either by logging in as one, or by prefacing the command with `sudo`. Please check that the installation script has executable permissions before trying to execute it.



Figure 2.2: Enter your password.

Next, you will be asked where to install the server (figure 2.3). If you do not have a particular reason to change this, simply leave it at the default setting. The chosen directory will be referred to as the *server installation directory* throughout the rest of this manual.

The installer allows you to specify the maximum amount of memory the CLC Server will be able to utilize (figure 2.4). The range of choice depends on the amount of memory installed on your system and on the type of machine used. If you do not have a reason to change this value you should simply leave it at the default setting.

If you are installing the Server on a Linux or Mac system, you are offered the option to specify a user account that will be used to run the *CLC Bioinformatics Database* process. Having a specific, non-root user for this purpose is generally recommended. On a standard setup, this would have the effect of adding this username to the service scripts, which can then be used for starting up and shutting down the *CLC Bioinformatics Database* service and setting the ownership of the files in the installation area. Downstream, the user running the *CLC Bioinformatics Database* process will own files created in File Locations, for example, after data import or data analyses.

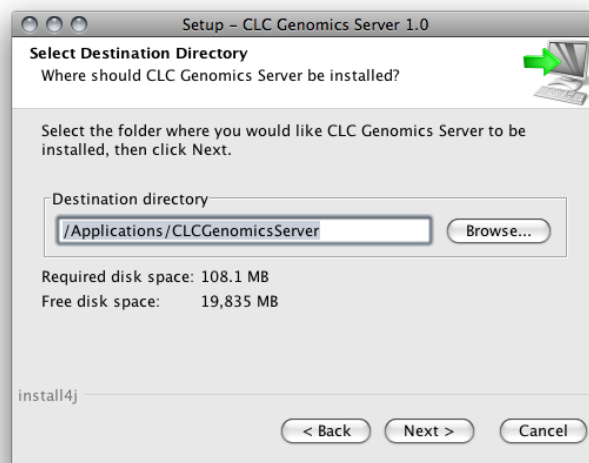


Figure 2.3: Choose where to install the server. Exemplified here with **CLC Genomics Server**

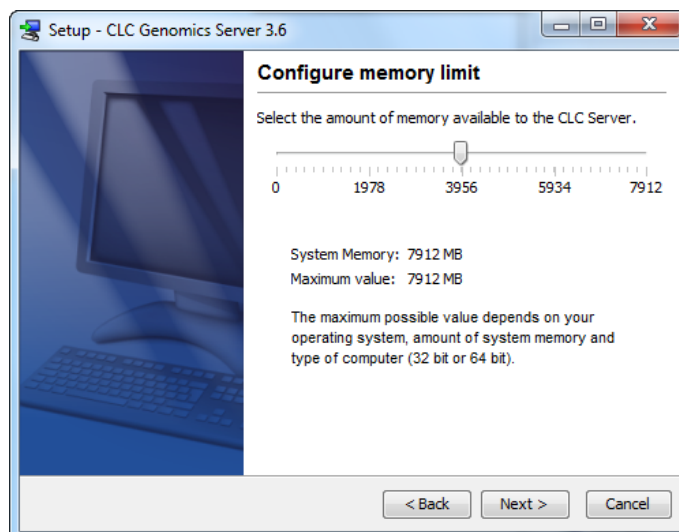


Figure 2.4: Choose the maximum amount of memory used by the server.

If you are installing the server on a Windows system you will be able to choose if the service is started manually or automatically by the system.

The installer will now extract the necessary files.

On a Windows system, if you have chosen that the service should be started automatically, the service should also start running at this point. On Linux or Mac, if you have chosen the option to start the system at the end of installation, the service should also have started running. Please note that if you do not already have a license file installed, then the *CLC Bioinformatics Database* process will be running in a limited capacity at this point. Downloading a license is described in section 2.7.

Information on stopping and starting the *CLC Bioinformatics Database* service is provided in section 2.8.

2.4 Installation modes - console and silent

Two installation modes are available to support efficient installation of the Workbench software.

- **Console mode** This mode is particularly useful when installing Workbenches onto remote systems. On Linux, this mode is enabled by using the option `-c` when launching the installer from the command line. On Windows the option is `-console`.
- **Silent mode** This mode supports hands-off installation. Default answers to all prompts are used, although a non-default installation directory can be specified if desired (see below). Silent mode is activated using the `-q` parameter when launching the installer from the command line. On Windows, the `-console` option can be appended after `-q`, that is, as *the second parameter*, to ensure output to the console.

If desired, you can **specify the directory to install the software to** when running the installer in silent mode. Do this adding the `-dir` option to the command line.

On Windows, the `-console` and the `-dir` options only work when the installer is run in silent mode.

The following is an example of a command that would install a Workbench into the directory "c:\bioinformatics\clc" on a Windows system using silent mode with console output :

```
CLCMainWorkbench_8_0.exe -q -console -dir "c:\bioinformatics\clc"
```

On a Linux system, a similar command to install into the directory `/opt/clcgenomicsworkbench11` could look like:

```
./CLCGenomicsWorkbench_11_0_1_64.sh -c -q -dir /opt/clcgenomicsworkbench11
```

The `-q` and the `-console` options work for the uninstall program as well.

2.5 Upgrading an existing installation

For a single *CLC Server*, the steps we recommend when upgrading to a new version are:

- Stop the *CLC Server* service after making sure that nobody is using the server. Mechanisms to help with this, including sending a message to users logged into the *CLC Server*, can be found in section 7.3. Getting information about who is logged in is described in section 7.1.
- Install the *CLC Server* software in the same directory the existing version was installed in. All settings will be maintained, for example, the locations data are stored, Import/Export directories, BLAST database locations, Users and Groups, and External Application settings.

When upgrading between major versions, there are extra steps to be taken. These are described in section 2.5.1). Major version lines are denoted by the first number in the version. For example, upgrading from software with version 10.0 to version 11.0 involves an upgrade to a new major version line.

2.5.1 Upgrading between major versions

There are a few extra steps to take beyond those outlined in section 2.5 when upgrading to a new major version line.

- An updated license file needs to be downloaded (see section 2.7), and the service restarted.
- All users of client software (CLC Workbenches and the CLC Server Command Line Tools) must upgrade their software. Corresponding and compatible software versions are listed at the bottom of the Latest Improvement listings for a given server version. e.g. for the latest release, this can be found at: <http://www.qiagenbioinformatics.com/products/clc-genomics-server/latest-improvements/current-line/>.

2.6 Allowing access through your firewall

By default, the server listens for TCP-connections on port 7777 (see section 3.3 for info about changing this).

If you are running a firewall on your server system you will have to allow incoming TCP-connections on this port before your clients can contact the server from a Workbench or web browser. Consult the documentation of your firewall for information on how to do this.

Besides the public port described above the server also uses an internal port on 7776. There is no need to allow incoming connections from client machines to this port.

2.7 Downloading a license

The *CLC Bioinformatics Database* looks for licenses in the `licenses` folder in the installation area. Downloading and installing licenses is similar for all supported platforms, but varies in certain details. Please check the platform-specific instructions below for how to download a license file on the system you are running the *CLC Bioinformatics Database* on or the section on downloading a license to a non-networked machine if the *CLC Bioinformatics Database* is running on a machine without a direct connection to the external network.

2.7.1 Windows license download

License files are downloaded using the `licensedownload.bat` script. To run the script, right-click on the file and choose **Run as administrator**. This will present a window as shown in figure 2.5.

Paste the Order ID supplied by QIAGEN (right-click to **Paste**) and press Enter. Please contact ts-bioinformatics@qiagen.com if you have not received an Order ID.

Note that if you are *upgrading* an existing license file, this needs to be deleted from the `licenses` folder. When you run the `downloadlicense.command` script, it will create a new license file.

Restart the server for the new license to take effect (see how to restart the server in section 2.8.1).

```

C:\WINDOWS\system32\cmd.exe
#####
##          CLC License Download Utility          ##
##          For the CLC License Server            ##
#####
Hostname:      machine hostname is printed here
Host ID(s):    machine hostids are printed here

Please enter (or copy/paste) your license order ID and press return...
>

```

Figure 2.5: Download a license based on the Order ID.

2.7.2 macOS license download

License files are downloaded using the `downloadlicense.command` script. To run the script, double-click on the file. This will present a window as shown in figure 2.6.

```

#####
##          CLC License Download Utility          ##
##          For the CLC License Server            ##
#####
Hostname:      machinehostname is printed here
Host ID(s):    machinehostids are printed here

Please enter (or copy/paste) your license order ID and press return...
>

```

Figure 2.6: Download a license based on the Order ID.

Paste the Order ID supplied by QIAGEN and press Enter. Please contact ts-bioinformatics@qiagen.com if you have not received an Order ID.

Note that if you are *upgrading* an existing license file, this needs to be deleted from the `licenses` folder. When you run the `downloadlicense.command` script, it will create a new license file.

Restart the server for the new license to take effect (see how to restart the server in section 2.8.2).

2.7.3 Linux license download

License files are downloaded using the `downloadlicense` script. Run the script and paste the Order ID supplied by QIAGEN Aarhus. Please contact ts-bioinformatics@qiagen.com if you have not received an Order ID.

Note that if you are *upgrading* an existing license file, this needs to be deleted from the `licenses` folder. When you run the `downloadlicense` script, it will create a new license file.

Restart the server for the new license to take effect (see how to restart the server in section 2.8.3).

2.7.4 Download a static license on a non-networked machine

To download a static license for a machine that does not have direct access to the external network, you can follow the steps below after the Server software has been installed.

- Determine the host ID of the machine the server will be running on by running the same

tool that would allow you to download a static license on a networked machine. The name of this tool depends on the system you are working on:

- Linux: downloadlicense
- Mac: downloadlicense.command
- Windows: licensedownload.bat

When you run the license download tool, the host ID for the machine you are working on will be printed to the terminal.

In the case of a job node setup, the only machine you need the host ID for is the master node. This is the machine the license file will be stored on.

- Make a copy of this host ID such that you can use it on a machine that has internet access.
- Go to a computer with internet access, open a browser window and go to the relevant network license download web page:

For the Genomics Server version 5.0 or higher, please go to:

<https://secure.clcbio.com/LmxWSv3/GetServerLicenseFile>

For the Genomics Server version 4.5.2 and lower, please go to:

<http://licensing.clcbio.com/LmxWSv2/GetServerLicenseFile>

It is vital that you choose the license download page appropriate to the version of the software you plan to run.

- Paste in your license order ID and the host ID that you noted down earlier into the relevant boxes on the webpage.
- Click on 'download license' and save the resulting .lic file.
- Take this file to the machine acting as the CLC Server master node and place it in the folder called 'licenses' in the CLC Server installation directory.
- Restart the CLC Server software.

2.8 Starting and stopping the server

2.8.1 Microsoft Windows

On Windows based systems the *CLC Bioinformatics Database* can be controlled through the Services control panel.

The *CLC Bioinformatics Database* service is called `CLCBioinformaticsDatabase`.

Choose the service and click the start, stop or restart link as shown in figure 2.7.

Once your server is started, you can use the Admin tab on the server web interface to manage your server operation (see section 7).

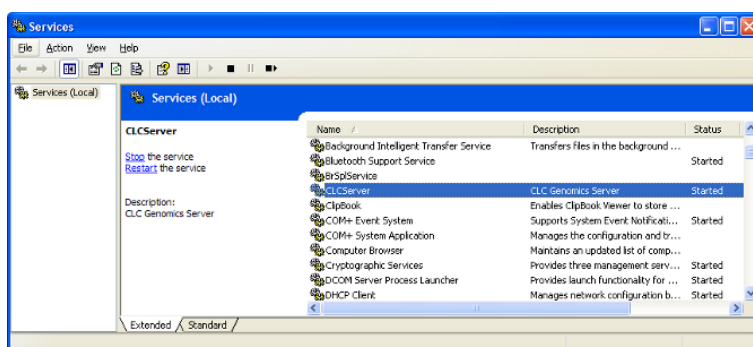


Figure 2.7: Stopping and restarting the server on Windows by clicking the blue links.

2.8.2 macOS

On macOS the server can be started and stopped from the command line.

Open a terminal and navigate to the *CLC Bioinformatics Database* installation directory. Once there, the server can be controlled with the following commands.

To start the server run the command:

```
sudo ./CLCBioinformaticsDatabase start
```

To stop the server run the command:

```
sudo ./CLCBioinformaticsDatabase stop
```

To view the current status of the server run the command:

```
sudo ./CLCBioinformaticsDatabase status
```

You will need to set this up as a service if you wish it to be run that way. Please refer to your operating system documentation if you are not sure how to do this.

Once your server is started, you can use the Admin tab on the server web interface to manage your server operation (see section 7).

2.8.3 Linux

You can start and stop the *CLC Bioinformatics Database* service from the command line. You can also configure the service to start up automatically after the server machine is rebooted.

During installation of the *CLC Bioinformatics Database* a service script is placed in `/etc/init.d/`.

This script will have a name reflecting the server solution, and it includes the name of the custom user account specified during installation for running the *CLC Bioinformatics Database* process.

Starting and stopping the service using the command line:

To start the *CLC Bioinformatics Database*:

```
sudo service CLCGenomicsServer start
```

To stop the *CLC Bioinformatics Database*:

```
sudo service CLCGenomicsServer stop
```

To restart the *CLC Bioinformatics Database*:

```
sudo service CLCGenomicsServer restart
```

To view the status of the *CLC Bioinformatics Database*:

```
sudo service CLCGenomicsServer status
```

Start service on boot up:

On Red Hat Enterprise Linux and SuSE this can be done using the command:

```
sudo chkconfig CLCGenomicsServer on
```

How to configure a service to automatically start on reboot depends on the specific Linux distribution. Please refer to your system documentation for further details.

Troubleshooting

If the *CLC Bioinformatics Database* is run as a service as suggested above, then the files in the installation area of the software and the data files created after installation in CLC Server File Locations will be owned by the user specified to run the *CLC Bioinformatics Database* process. If someone starts up the *CLC Bioinformatics Database* process as root (i.e. an account with super-user privileges) then the following steps are recommended to rectify the situation:

1. Stop the *CLC Bioinformatics Database* process using the script located within the installation area of the *CLC Bioinformatics Database* software. You can do that using the full path to this script, or by navigating to the installation area and running:

```
sudo ./CLCGenomicsServer stop
```

2. Change ownership recursively on all files in the installation area of the software and on all areas specified as Server File Locations.
3. Start the *CLC Bioinformatics Database* service as the specified user by using the service script:

```
sudo service CLCGenomicsServer start
```

4. In case the server still fails to start correctly it can be started in the foreground with output being written to the console to help identify the problem. It is done by running:

```
sudo ./CLCGenomicsServer start-launchd
```

Once your server is started, you can use the Admin tab of the web administrative interface to manage your server operation (see section 7).

Chapter 3

Configuration

3.1 Logging into the administrative interface

The administrative interface for a running *CLC Bioinformatics Database* is accessed via a web browser. Most configuration occurs via this interface. Simply type the host name of the server machine you have installed the *CLC Bioinformatics Database* software on, followed by the port it is listening on. Unless you change it, the port number is 7777. An example would be

```
http://clccomputer:7777/ or http://localhost:7777/
```

The default administrative user credentials are:

- **User name:** root
- **Password:** default

Use these details the first time you log in. We recommend that you change this password.

Details of how to change the administrative user password is covered in section [4.1](#).

3.2 Adding locations for saving data

One or more locations for storing your data must be configured before the system can be used.

Setting up an sql database to be used to store CLC data is described in (section [3.2.3](#)).

File system locations can also be configured, as described in (section [3.2.1](#)).

3.2.1 File system locations

Configurations relating to data storage on the *CLC Bioinformatics Database* are available via the web interface when logged in as a user with administrative privileges. Navigate to the *Admin* tab, click on the *Main configuration* tab, and then click on the **File system locations** heading to open that area if it is not already open.

Adding a file system location

To add a file system location for storing CLC data, click on the **Add New File Location** button. See figure 3.1.

The screenshot shows the 'Main configuration' window with the 'Admin' tab selected. Under 'File system locations', there are two entries. Each entry consists of a checked 'Path' checkbox, a text input field containing a path, and an unchecked 'Permissions enabled' checkbox. To the right of each path field are 'Remove Location' and 'Rebuild Index' buttons. Below the locations is the 'Data compression' section, which has a checked checkbox for 'Save CLC data elements in a compressed format.' and a link to the manual. At the bottom, there are four expandable sections: 'Import/export directories', 'Automatic recycle bin cleanup', 'Direct data transfer from client systems', and 'HTTP settings'. A 'Save Configuration' button is at the very bottom.

Figure 3.1: File system location settings. The checkmark to the left of a configured location indicates it is available for use by those logged into the server. Internal data compression is enabled by default. This setting applies to all configured file system locations.

In this dialog, enter the path to the folder where data imported into or created by the *CLC Bioinformatics Database* will be stored. The path should point to an *existing* folder on the server machine, and the user *running the server process* needs to have read and write access to the folder. This is usually a dedicated user, or it may be the system's root user if you have not created a dedicated user for this purpose.

The checkbox to the left of each location indicates whether or not that location should be available to users. By default, this option is enabled, meaning the location is accessible to those connecting to the *CLC Bioinformatics Database* using CLC Workbenches or the CLC Server Command Line Tools. Each enabled location will appear in the Workbench **Navigation Area** the next time it is connected to the server. Unchecking this box and saving the configuration makes the location unavailable for use.

By default, every enabled location is available to any user who can log into the *CLC Bioinformatics Database*. However, access control can be configured, if desired. The first step is to check the **Permissions enabled** box for each location where access controls will be applied. Further configuration details are described in section 5.1.

Rebuilding the index is described in section 3.2.4. There is no need to do this when adding a new area as a file system location.

Clicking on the **Remove Location** button removes the location from the list that can be used by the *CLC Bioinformatics Database*. The underlying folder and its contents are **not** deleted. The folder, and the data it contains, can be made accessible again simply by going through the same steps outlined above.

Once you have pressed **Save Configuration**, this location will be added and it should now appear in the list at the left hand side of the web client.

The **Data compression** option in figure 3.1 is described in section 3.2.2.

Important points about CLC Server file system locations

- Folders added as file system locations should be **dedicated for use** by the *CLC Bioinformatics Database* and should be directly accessed only by the *CLC Bioinformatics Database*.
- The underlying file system must support file locking.
- All the data written to file system locations by the *CLC Bioinformatics Database* will be in clc format and will be owned by the user that runs the *CLC Bioinformatics Database* process.
- Files should **not** be moved manually into folders designated as CLC Server file system locations, or their subfolders, for example using standard operating system's command tools, drag and drop, and so on.
- Areas designated as file system locations should **not** overlap. That is, if a folder has been designated as a file system location, it should not be a subfolder of another area also designated as a file system location, or vice versa. Overlapping file system locations lead to problems with data indexing, which in turn leads to problems finding the stored data. Further information about indexing can be found in section 3.2.4.

3.2.2 Enabling and disabling internal compression of CLC data

CLC data stored using internal compression takes less space. An option is provided under the **Data compression** heading to turn off internal data compression. See figure 3.1. Enabling data compression may impose a performance penalty depending on the characteristics of the hardware used. This penalty is typically small, and we generally recommend that this option remains enabled.

This setting applies to all configured file system locations. Any change will apply to data imported into or created after the change is made. Existing data is not affected.

Internal data compression and data compatibility Internal compression was introduced with the CLC Genomics Server 11.0 and corresponding client software, the CLC Genomics Workbench 12.0 and CLC Main Workbench 8.1. Data imported into or created by these systems and newer versions are compressed by default. However, software older than this, including retired products, cannot read the compressed data format.

Data created using older versions of CLC software will not be stored in compressed format.

To facilitate sharing particular datasets with people using older versions of the software, an option is available when exporting the data to CLC or zip format to export without this compression.

3.2.3 Database locations

Before adding a database location, you need to set up the database itself. This is described in section 2.2.

To set up a database location on the CLC Server,

- Open a web browser and navigate to the web administrative interface.
- Go to the *Admin* tab and open the *Main configuration* section.
- Under the **Database locations** heading, click the **Add New Database Location** button. This pops a window like that shown in figure 3.2.
- To configure the database location, enter the host and port information and select the database type. The *Database type* drop down list contains the types for which drivers are available. A connection string is generated from this. A custom connection string can be entered instead. Add the user name and password information for the user role on your Database Management System (DBMS), see section 2.2.

Figure 3.2: Adding a new database location. Here, two drivers are available to the CLC Server, a MySQL driver and an Oracle driver.

If an Oracle database driver is available to the *CLC Bioinformatics Database*, two items will be presented in the Database type drop down list, as shown in figure 3.2. The one shown as "Oracle" is the traditional one, which uses the SID style (e.g. `jdbc:oracle:thin:@[HOST][:PORT]:SID`). The other, "Oracle Service", uses the thin-style service name (e.g. `jdbc:oracle:thin:@//[HOST][:PORT]/SERVICE`).

- Click the *Save Configuration* button to save the configuration.

The newly added database location should now appear in the **Navigation Area** in the left hand side of the window.

3.2.4 Rebuilding the index

The server maintains an index of all the elements in the data locations. The index is used when searching for data. For all locations you can choose to **Rebuild Index**. This should be done only when a new location is added or if you experience problems while searching (e.g. something is missing from the search results). This operation can take a long time depending on how much data is stored in this location.

If you move the server from one computer to another, you need to move the index as well. Alternatively, you can re-build the index on the new server (this is the default option when you add a location). If the rebuild index operation takes too long and you would prefer to move the old index, simply copy the folder called `searchindex` from the old server installation folder to the new server.

The status of the index server can be seen in the **User Statistics** pane found in the **Status and Management** tab page showing information on where the index server resides and the number of locations currently being serviced.

3.3 Changing the listening port

The default listening port for the CLC Server is 7777. This has been chosen to minimize the risk of collisions with existing web-servers using the more familiar ports 80 and 8080. If you would like to have the server listening on port 80 in order to simplify the URL, this can be done in the following way.

- Navigate to the CLC Server installation directory.
- Locate the file called `server.xml` in the `conf` directory.
- Open the file in a text editor and locate the following section

```
<Connector port="7777" protocol="HTTP/1.1"
  connectionTimeout="20000"
  redirectPort="8443" />
```

- Change the port value to desired listening port (80 in the example below)

```
<Connector port="80" protocol="HTTP/1.1"
  connectionTimeout="20000"
  redirectPort="8443" />
```

- Restart the service for the change to take effect (see how to restart the server in section 2.8).
- Once the service is restarted, please log into the administrative interface and change the default port number in the "Master node port" field under Admin | Job distribution | Server setup, then click on **Save Configuration** button to save the new setting.

3.4 Setting the amount of memory available for the JVM

When running the *CLC Bioinformatics Database*, the Java Virtual Machine (JVM) needs to know how much memory it can use. This depends on the amount of physical memory (RAM) and can

thus be different from computer to computer. Therefore, the installer investigates the amount of RAM during installation and sets the amount of memory that the JVM can use.

On **Windows** and **Linux**, this value is stored in a property file called `ServerType.vmoptions` (e.g. `CLCGenomicsServer.vmoptions`) which contains a text like this:


```
-Xmx8192m
```

The number (8192) is the amount of memory in megabytes the *CLC Bioinformatics Database* is allowed to use. This file is located in the installation folder of the *CLC Bioinformatics Database* software.

By default, the value is set to 50% of the available RAM on the system you have installed the software on.

You can manually change the number contained in the relevant line of the `vmoptions` file for your *CLC Bioinformatics Database* if you wish to raise or lower the amount of RAM allocated to the Java Virtual Machine.

3.5 HTTP settings

Under the **Admin**  tab, click **Configuration**, and you will be able to specify HTTP settings. Here you can set the time out for the user HTTP session and the maximum upload size (when uploading files through the web interface).

3.6 Deployment of server information to CLC Workbenches

See the *Deployment manual* at <http://www.qiagenbioinformatics.com/support/manuals/> for information on pre-configuring the server log-in information when Workbench users log in for the first time.

Chapter 4

Managing users and groups

4.1 Logging in the first time and changing the root password

When the server is installed, you will be able to log in via the web interface using the following credentials:

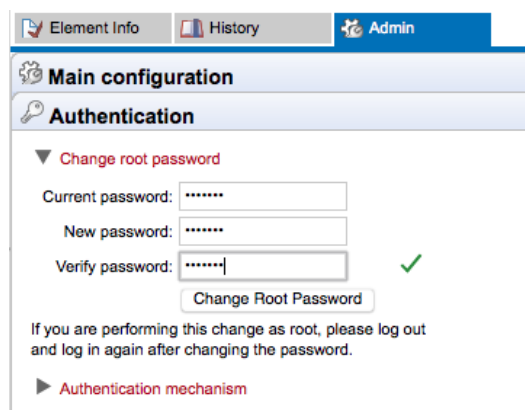
- **User name:** `root`
- **Password:** `default`

Once logged in, you should as a minimum set up user authentication (see section 4.2) and data locations (see section 3.2) before you can start using the server.

For security reasons, you should change the root password (see figure 4.1):

Admin (🔧) | Authentication (🔑) Change root password

Note that if you are going to use job nodes, it makes sense to set these up before changing the authentication mechanism and root password (see section ??).



The screenshot shows a web interface with a top navigation bar containing 'Element Info', 'History', and 'Admin' (highlighted in blue). Below the navigation bar is a 'Main configuration' section with a sub-section 'Authentication' (indicated by a key icon). Under 'Authentication', there is a collapsible section 'Change root password' (indicated by a downward arrow). This section contains three password input fields: 'Current password:', 'New password:', and 'Verify password:'. The 'Verify password:' field has a green checkmark to its right. Below the input fields is a button labeled 'Change Root Password'. A message below the button reads: 'If you are performing this change as root, please log out and log in again after changing the password.' At the bottom of the 'Authentication' section, there is a link for 'Authentication mechanism'.

Figure 4.1: We recommend changing the root password. The verification of the root password is shown with the green checkmark.

4.2 User authentication using the web interface

When the server is installed, you can log in using the default root password (username=root, password=default).

Once logged in, you can specify which of the three modes of authentication should be used by going to:

Admin (⚙️) | Authentication (🔑) Authentication mechanism

The three different modes of authentication are shown in figure 4.2.

If LDAP or Active Directory are selected, a settings panel is revealed, where the details of the integration are specified. An example for LDAP settings is shown in figure 4.3.

Members of a group specified as an administrative group with login rights to the *CLC Bioinformatics Database* will be configure the *CLC Bioinformatics Database* using the functionality under the Admin tab of the web administrative interface, as well as set permissions on folders of data, as described in section 5. For the built-in authentication method, this means adding particular users to the built-in **admin** group. For Active Directory or LDAP, this means designating a group in the box labeled **Admin group name** and adding any users who should be administrators of the CLC Server to this group.

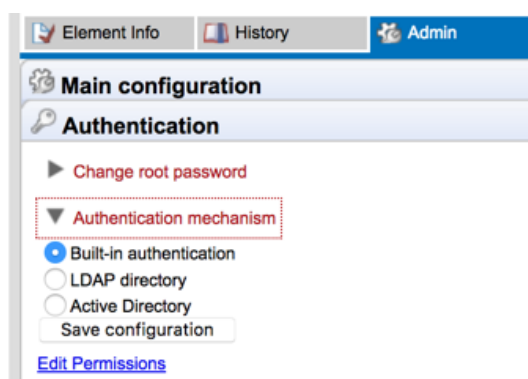


Figure 4.2: Three modes of user authentication are available. Clicking on the *Edit Permissions* link at the bottom opens up the *Global Permissions* tab, where access to the server and its functionality can be configured, as described in section ??.

4.2.1 Authentication options

Built-in authentication

This option will enable you to set up user authentication using the server's built-in user management system. This means that you create users, set passwords, assign users to groups and manage groups using the web interface (see section 4.2.2) or using the Workbench (see section 4.2.4). All the user information is stored on the server and is not accessible from other systems.

LDAP directory

This option will allow you to use an existing LDAP directory. This means that all information needed during authentication and group memberships is retrieved from the LDAP directory (figure 4.3).

Authentication

▼ Authentication mechanism

☐ Built-in authentication

☒ LDAP directory

☐ Active Directory

Hostname: host.example.com

Port: 389 Default if "ldaps://" is selected: 636. Else: 389

Encryption: ☒ Plain text Default: "Plain text"

☐ Forced Start TLS

☐ ldaps://

Disable SSL certificate check: ☐

Base DN: dc=example,dc=com

Admin group name: Default: admins

Cache timeout: Default: 3600 (seconds)

Users DN: ou=users (Base DN will be appended)

Groups DN: ou=groups (Base DN will be appended)

UID attribute: Default: uid

Group name attribute: Default: cn

Membership attribute: Default: memberUid

Bind DN: Leave empty to use anonymous bind for the initial lookup, used to get a user DN

Bind password:

DN to use for lookups: ☒ User DN Select the DN to be used for all search and read operations except for the initial one, for example user and email lookups. The Bind DN option is enabled for selection when Bind DN and password details have been entered above.

☐ Bind DN

Kerberos/GSSAPI Authentication: ☐

Kerberos realm: Leave empty to use default realm

Kerberos config file: Default: /etc/krb5.conf

Figure 4.3: LDAP settings panel.

If needed, the LDAP integration can use Kerberos/GSSAPI. Encryption options (Start TLS and LDAP over SSL) are available. If your LDAP server uses a certificate that is not generally trusted by the server system that the CLC Server software is running on, then it must be added to the truststore of the CLC Server installation (`CLC_SERVER_BASE/jre/lib/security/cacerts`, where `CLC_SERVER_BASE` is the server installations root location). This can be done with the `keytool` shipped with Java installations (also available in the `CLC_SERVER_BASE/jre/bin/keytool`), with a command like:

```
CLC_SERVER_BASE/jre/bin/keytool -import -alias \
  ldap_certificate -file LDAP_CERTIFICATE.cer -keystore \
  CLC_SERVER_BASE/jre/lib/security/cacerts -storepass changeit
```

Replace `LDAP_CERTIFICATE` with the path to the certificate your LDAP server uses for Start TLS/LDAPS connections. Replace `CLC_SERVER_BASE` with the path to the servers installation location.

For a node setup, this must be done for all job nodes as well.

Caution: If you update the server installation or reinstall the server, all imported certificates will be removed, and have to be imported again. You should also be aware that certificates have an expiration date, and will not be trusted after this date. Make sure to add a new certificate in advance of the expiration date.

The **DN to use for lookups** configuration allows you to choose which bind should be used for read

and search operations. If no bind DN have been entered an unauthenticated bind will be used to do the initial lookup (lookup of users DN based on the username), and all other read and search operations will be performed with users binds. If the **Bind DN** and **Bind password** have been filled in, you have the choice between using the 'Bind DN' or the 'User DN' for read and search operations, the 'Bind DN' will in this case always be used for the initial lookup.

Active Directory

This option will allow you to use an existing Active directory. This means that all information needed during authentication and group memberships is retrieved from the Active directory. Encryption options (Start TLS and LDAP over SSL) are available. Please see the notes about certificates in the LDAP section (see section 4.2.1) above for details.

4.2.2 Managing users and groups using built-in authentication

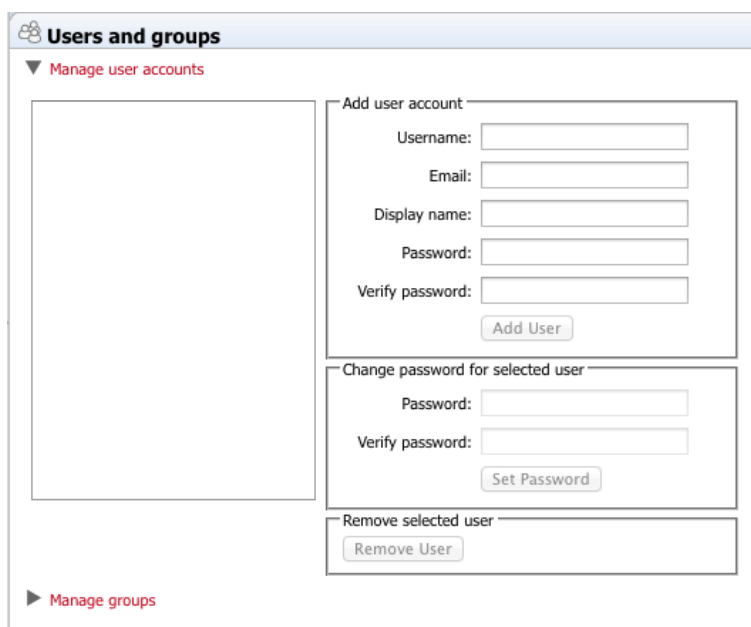
This information only applies if built-in authentication is being used.

Managing users via the web administrative interface

To create or remove users or change their password:

Admin (🔑) | Users and groups (👤) Manage user accounts

This will display the panel shown in figure 4.4.



The screenshot shows a web interface titled 'Users and groups' with a sub-section 'Manage user accounts'. On the left is a large empty box for a user list. On the right are three form panels: 'Add user account' with fields for Username, Email, Display name, Password, and Verify password, plus an 'Add User' button; 'Change password for selected user' with fields for Password and Verify password, plus a 'Set Password' button; and 'Remove selected user' with a 'Remove User' button. At the bottom left, there is a link to 'Manage groups'.

Figure 4.4: Managing users.

Managing groups via the web administrative interface

To create or remove groups or change group membership for users when using built-in authentication, go to:

Admin (🔑) | Users and groups (👤) Manage groups

This will display the panel shown in figure 4.5.

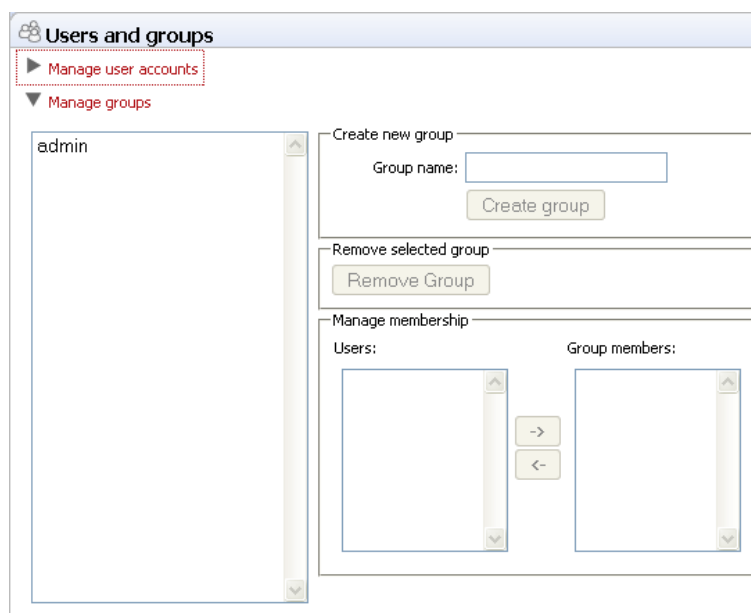


Figure 4.5: Managing users.

The same user can be a member of several groups.

Users who should have access to the administrative part of the server should be part of the "admin" group which is the only group with special permissions by default. The admin group already exists in a default setup of the *CLC Bioinformatics Database*.

You will always be able to log in as the *CLC Bioinformatics Database* root user, and this user has administrative level access rights.

4.2.3 User authentication using the Workbench

This information only applies if built-in authentication is being used. If LDAP or AD is being used, the menus described here will be disabled.

Users and groups can be managed through the Workbench by logging into the *CLC Bioinformatics Database* as an administrative user and then going to the Workbench menu:

File | Manage Users and Groups

This will display the dialog shown in figure 4.6.

4.2.4 Managing users through the Workbench

This information only applies if built-in authentication is being used. If LDAP or AD is being used, the menus described here will be disabled.

Click the **Add** (+) button to create a new user. Enter the name of the user and enter a password. You will be asked to re-type the password. If you wish to change the password at a later time, select the user in the list and click **Change password** (🔑).

To delete a user, select the user in the list and click **Delete** (✖).

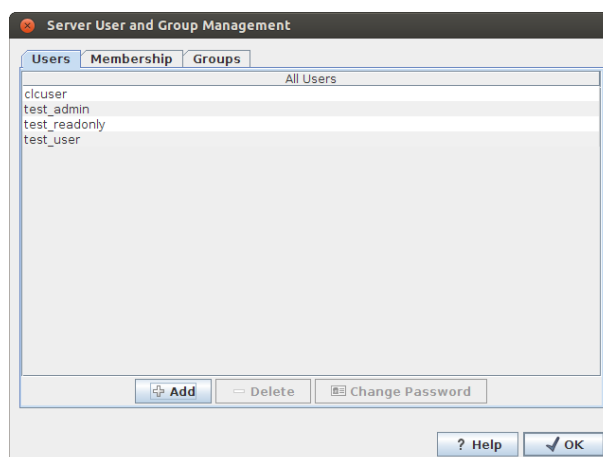


Figure 4.6: Managing users.

4.2.5 Managing groups through the Workbench

This information only applies if built-in authentication is being used. If LDAP or AD is being used, the menus described here will be disabled. Access rights are granted to groups, not users, so a user has to be a member of one or more groups to get access to the data location. Here you can see how to add and remove groups, and next you will see how to add users to a group.

Adding and removing groups is done in the **Groups** tab (see figure 4.7).

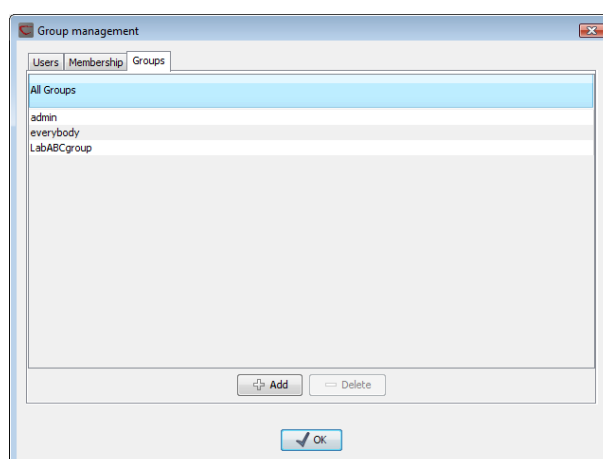


Figure 4.7: Managing groups.

To create a new group, click the **Add** (+) button and enter the name of the group. To delete a group, select the group in the list and click the **Delete** (=) button.

When a new group is created, it is empty. To assign users to a group, click the **Membership** tab. In the **Selected group** box, you can choose among all the groups that have been created. When you select a group, you will see its members in the list below (see figure 4.8). To the left you see a list of all users.

To add or remove users from a group, click the **Add** (➡) or **Remove** (⬅) buttons. To create new users, see section 4.2.4.

The same user can be a member of several groups.

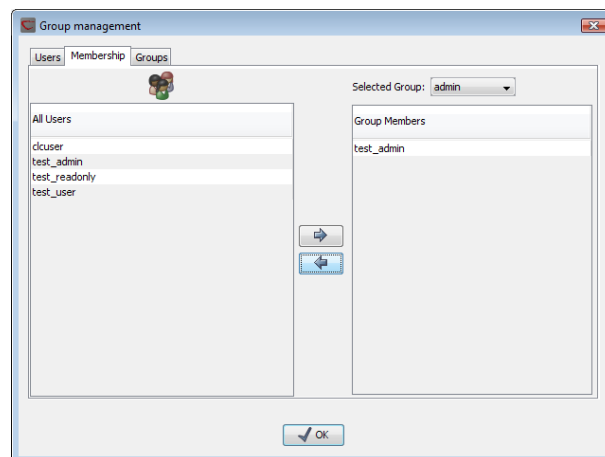


Figure 4.8: Listing members of a group.

Chapter 5

Access privileges and permissions


Server administrators can restrict access to members of specified groups at various levels:


- **Access to the CLC Server**, described in section ??.
- **Access to data in the server's file and data locations**, described in section 5.1.
- **Launching jobs** can be restricted to particular groups of users. Settings for data import and export are available under the Core tasks section.
- **Access to import/export directories**, described in section ??.

5.1 Controlling access to CLC Server data

The *CLC Bioinformatics Database* uses folders as the basic unit for controlling access to data, and access is granted (or denied) to groups of users.

Two types of access can be granted to a group on any folder within a server location:

Read access Users of the designated group(s) can see the elements in the folder, open them and copy from them. Access can be through any route, for example, via the CLC Command Line Tools or via the Workbench, for example when browsing in the **Navigation Area** of a Workbench, searching, or when clicking the "originates from" link in the **History** () of data.

Write access Users of the designated group(s) can make and **Save** () changes to an element, and new elements and subfolders can be created in that area.

For a user to be able to access a folder, they must have read access to all the folders above it in the hierarchy. In the example shown in figure 5.1, to access the *Sequences* folder, the user must have access to both the *Example Data* and *Protein* folders.

It is fine to just give write access to the final folder. For example, read access only could be granted to the *Example Data* and *Protein* folders, with read and write access granted to the *Sequences* folder.

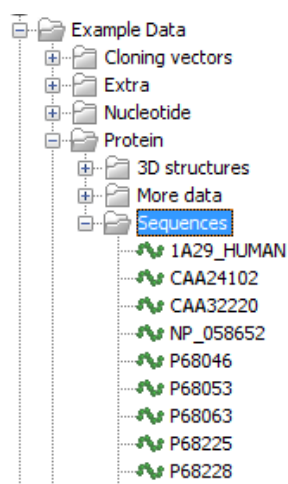


Figure 5.1: A folder hierarchy on the server.

Permissions on CLC Server File Locations must be **explicitly enabled** via the web administrative interface if they are desired. See section 3.2.1). Please see 5.1.3 for further details about the system behavior if permissions are not enabled and configured.

Configuring the permissions is done via a CLC Workbench acting as a client for the CLC Server. At the point when permissions are enabled on a File Location via the server web administrative interface, Only the *CLC Bioinformatics Database* root user or users in a configured admin group have access to data held in that File Location at this point. No groups will have read or write access to any area under this location. Permissions should then be explicitly set by the root or other admin user on the folders in that area, as described below.

5.1.1 Setting permissions on a folder

This step is done from within a CLC Workbench. Start up a copy of a CLC Workbench click on the menu option:

File | CLC Server Connection (S)

Log into the CLC Server as an administrative user.

You can then set permissions on folders within File Locations that have had permissions enabled, or on Database Locations, if you have a CLC Bioinformatics Database.

right-click the folder (F) | Permissions (P)

This will open the dialog shown in figure 5.2.

Set the relevant permissions for each of the groups and click **OK**.

If you wish to apply the permissions recursively, that is to all subfolders, check **Apply to all subfolders** in the dialog shown in figure 5.2. **Note** that this operation is usually only relevant if you wish to clean up the permission structure of the subfolders. **It should be applied with caution**, since it can potentially destroy valuable permission settings in the subfolder structure.

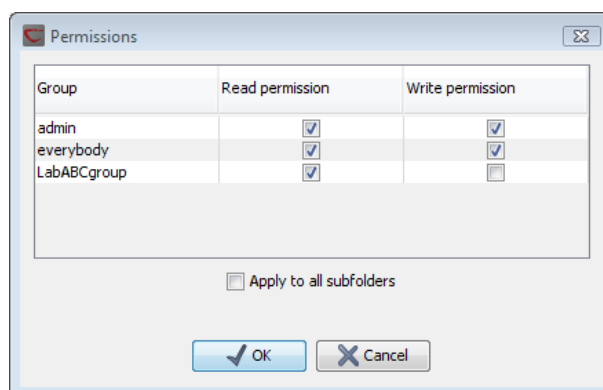


Figure 5.2: Setting permissions on a folder.

5.1.2 Recycle bin

When users delete data in the **Navigation Area** of the Workbench, it is placed in the recycle bin. When the data is situated on a data location on a *CLC Bioinformatics Database*, the data will be placed in a recycle bin for that data location. Each user has an individual recycle bin containing the data deleted by that particular user which cannot be accessed by any other user (except server administrators, see below). This means that any permissions applied to the data prior to deletion are no longer in effect, and it is not possible to grant other users permission to see it while it is located in the recycle bin. In summary, the recycle bin is a special concept that is not included in the permission control system.

Server administrators can access the recycle bins of other users through the Workbench:

right-click the data location (📁) | Location | Show All Recycle Bins

This will list all the recycle bins at the bottom of the location as shown in figure 5.3.

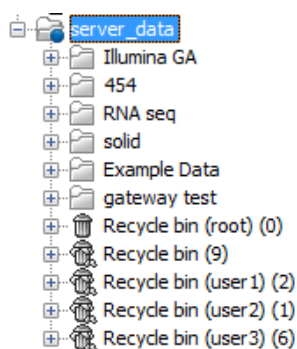


Figure 5.3: Showing all recycle bins.

The recycle bin without a name contains all the data that was deleted in previous versions of the *CLC Bioinformatics Database* before the concept of a per-user recycle bin was introduced. This recycle bin can only be accessed by server administrators by selecting **Show All Recycle Bins**.

The administrator is also able to empty the recycle bin of a user:

right-click the recycle bin (🗑️) | Empty

All recycle bins can be emptied in one go:

right-click the data location (📁) | Location | Empty All Recycle Bins

Please note that these operations cannot be undone.

CLC Bioinformatics Database can be set to automatically empty recycle bins when the data has been there for more than 100 days. This behavior can be controlled for each data location: Under the **Main configuration** heading, click the **Automatic recycle bin clean-up** header and click the **Configure** button. This will allow you to disable the automatic clean-up completely or specify when it should be performed as shown in figure 5.4.

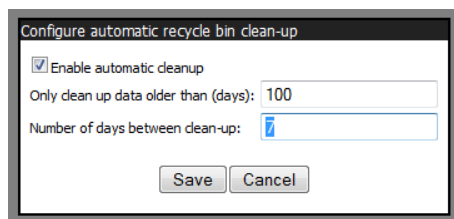


Figure 5.4: Automatic clean-up of the recycle bin.

Data deleted before the per-user recycle bin concept was introduced will be ignored by the automatic clean-up (this is the data located in the general recycle bin that is not labeled with a user name).

5.1.3 Technical notes about permissions and security

All data stored in *CLC Bioinformatics Database* file system locations are owned by the user that runs the *CLC Bioinformatics Database* process. Changing the ownership of the files using standard system tools is not recommended and will usually lead to serious problems with data indexing and hamper your work on the *CLC Bioinformatics Database*.

One implication of the above ownership setup is that by default, (i.e. without permissions enabled), all users logging into the *CLC Bioinformatics Database* are able to access all data within that file system location, and write data to that file system locations. All files created within such a file system location are then also accessible to all users of the *CLC Bioinformatics Database*.

Group permissions on file system locations is an additional layer within the *CLC Bioinformatics Database*, and is not part of your operating system's permission system. This means that enabling permissions, and setting access restrictions on CLC file system locations only affects users accessing data through CLC tools (e.g. using a Workbench, the CLC Command Line Tools, the *CLC Bioinformatics Database* web interface or the Server API). If users have direct access to the data, using for example general system tools, the permissions set on the data in *CLC Bioinformatics Database* has no effect.

Chapter 6

Queue

Clicking the **Queue** panel will show a list of all the processes that are currently in the queue including jobs in progress.

For each process, you are able to **Cancel** (☐) the processes. At the top, you can see the progress of the process that is currently running.

Chapter 7

Status and management

Server operation can be managed from the Admin tab, under Status and Management (figure 7.1).

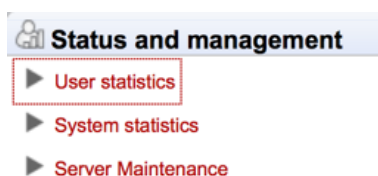


Figure 7.1: The Status and management tab.

7.1 User statistics

The User statistics section contains information about the number of users logged in, the number of active sessions (number of logins), and information about each active session. An example is shown in figure 7.2.

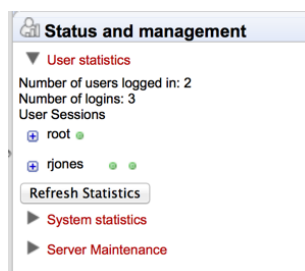


Figure 7.2: Information about the number of users and active sessions (logins) is provided in the User statistics area. Here, two users are logged in. rjones has two active sessions and root has one.

A green dot by a user's name indicates that they are logged into the server. Two green dots indicate that this user is logged in twice. For example, perhaps they have Workbenches running on two different systems and are logged in via both Workbenches. A grey dot means they have previously logged in but are not at this time.

Click on the small button with a plus to the left of a username to expand the information about that user's sessions (figure 7.3). You can also log users off the server by clicking on the **Invalidate Session...** button. This opens a confirmation dialog where a message to the user can

be written. This message is displayed via the user's active session. For example, if they are logged into a Workbench, a window will pop up saying they have been logged out of the server and also containing the message written in this field. This action forcibly logs the user out of the *CLC Bioinformatics Database*. This action does **not** stop jobs already submitted or running on the server. Optionally, you can send a message to the user whose session is being terminated (see figure 7.4). If the user is logged into the *CLC Bioinformatics Database* from a CLC Workbench, then the message entered will appear in a warning box that pops up via the CLC Workbench.



Figure 7.3: Details about *jbloggs*' session can be seen by clicking on the small button to the left of that username.

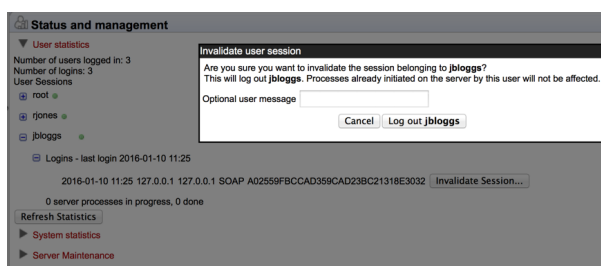


Figure 7.4: Clicking on the *Invalidate Session* button will forcibly log a user out of the CLC Server. The admin can optionally provide a message to the user when doing this.

7.2 System statistics

Crashed threads, suggesting system level problems, are reported in this area. In some instances, a system restart may be needed to resolve the issue.

The message "No system level problems detected" is shown in this area if no problems have been detected. An example of the information provided when a problem is detected is shown in figure 7.5. In the case shown, the job submission threads were dead, with the problem reported here and in more detail in the CLC Server log files.

7.3 Server maintenance

Settings under the Server maintenance tab allow a server administrator to change the operating mode of the server and send out messages to users of the *CLC Bioinformatics Database* (see figure 7.6).

- **Normal Operation** The *CLC Bioinformatics Database* is running.

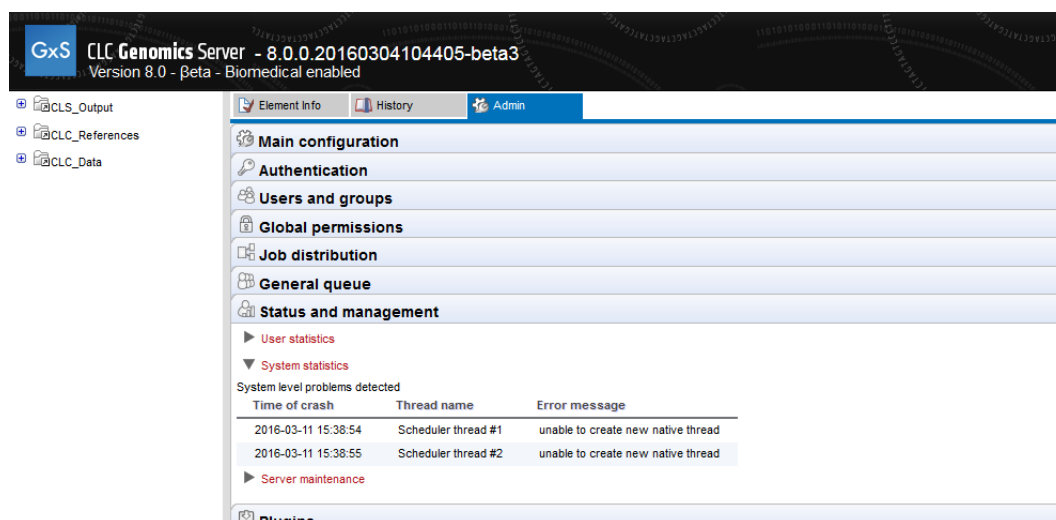


Figure 7.5: System level problems detected and reported in the system statistics area.

- **Maintenance Mode** Current jobs are allowed to run and complete, but submission of new jobs is restricted. While the server is in maintenance mode, users already logged in can check the progress of their jobs or view their data, but they cannot submit new jobs. Users not already logged in cannot log in. An administrator can write a warning message, for example, to inform users about the expected period of time the server will be in maintenance mode.
- **Log Out Users** All users currently logged in will be logged out. All running jobs will be allowed to complete. No users can log in while in this mode. An administrator can also write a warning message for the users.
- **Shut down** The *CLC Bioinformatics Database* and any attached job nodes will shut down.
- **Restart** The *CLC Bioinformatics Database* and any attached job nodes will be shut down and restarted.

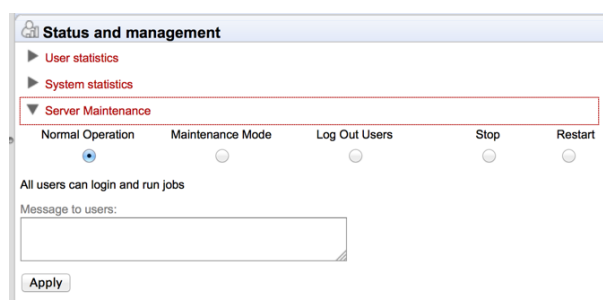


Figure 7.6: The server administrator can control the operating mode of the CLC Server from under the Server maintenance tab.

Chapter 8

Audit log

The audit log records actions performed on the *CLC Bioinformatics Database*. Included are actions like logging in, logging out and import. Data management operations such as copying, deleting and adding files are not Server actions and are thus not recorded.

Audit log information is available via the web administrative interface under the **Audit log** tab. The information presented here is stored in a database. Once a month, and when the *CLC Bioinformatics Database* is started up, entries in the audit log older than 3 months are deleted.

The limit the audit log database can grow to is 64 GB. If a new entry will push the size past this limit, the system will remove some of the oldest entries so that it is possible for newer entries to be added.

Audit information is also written to text-based log files. Upon the first activity on a given date, a new log file called `audit.log` is created. This file is then used for logging that activity and subsequent Server activities on that day. When this new `audit.log` file is created, the file that previously had that name is renamed to `audit.<actual events date>.log`. These log files are retained for 31 days. When the creation of a new `audit.log` file is triggered, audit log files older than 31 days are checked for and deleted.

The audit log files can be found under the Server installation area under `webapps/CLCServer/WEB-INF`.

The audit log text files are tab delimited and have the following fields:

- Date and time
- Log level
- Operation: Login, Logout, Command queued, Command done, Command executing, Change server configuration, Server lifecycle; more may be added and existing may be changed or removed.
- Users
- IP Address
- Process name (when operation is one of the Command values) or description of server lifecycle (when operation is Server lifecycle)

- Process identifier - can be used to differentiate several processes of the same type.
- Status - can be used to identify whether the entry was successful or not, e.g. if a job execution failed it will be marked here. Any number other than 0 means failed.

8.1 Customized attributes on data locations

Location-specific attributes can be set on all elements stored in a given data location. Attributes could be things like company-specific information such as LIMS id, freezer position etc. Attributes are set using a CLC Workbench acting as a client to the CLC Server.

Note that the attributes scheme belongs to a particular data location, so if there are multiple data locations, each will have its own set of attributes.

To configure which fields that should be available¹ go to the Workbench:

right-click the data location | Location | Attribute Manager

This will display the dialog shown in figure 8.1.

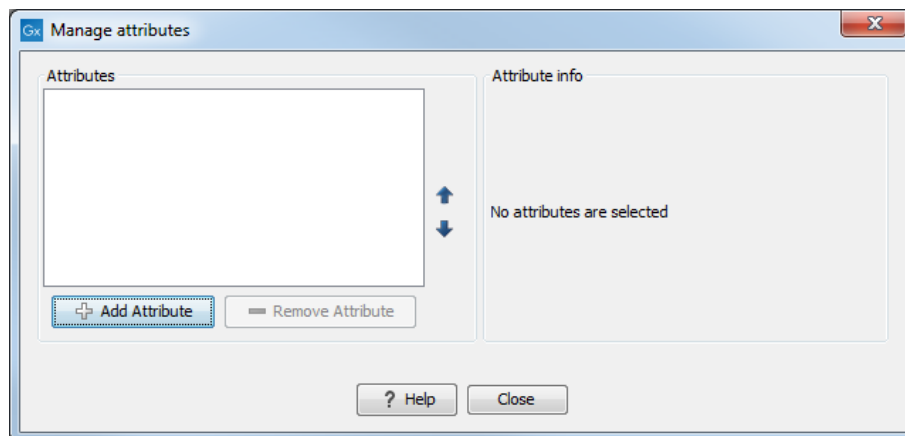


Figure 8.1: Adding attributes.

Click the **Add Attribute** (+) button to create a new attribute. This will display the dialog shown in figure 8.2.

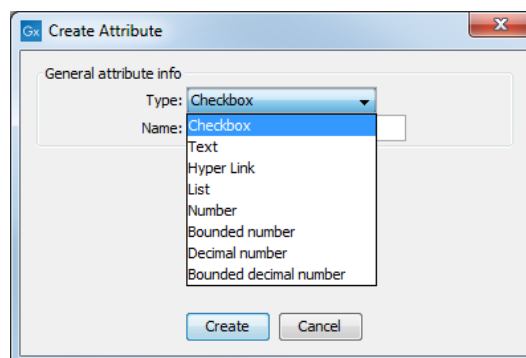


Figure 8.2: The list of attribute types.

First, select what kind of attribute you wish to create. This affects the type of information that

¹If the data location is a server location, you need to be a server administrator to do this.

can be entered by the end users, and it also affects the way the data can be searched. The following types are available:

- **Checkbox.** This is used for attributes that are binary (e.g. true/false, checked/unchecked and yes/no).
- **Text.** For simple text with no constraints on what can be entered.
- **Hyper Link.** This can be used if the attribute is a reference to a web page. A value of this type will appear to the end user as a hyper link that can be clicked. Note that this attribute can only contain one hyper link. If you need more, you will have to create additional attributes.
- **List.** Lets you define a list of items that can be selected (explained in further detail below).
- **Number.** Any positive or negative integer.
- **Bounded number.** Same as number, but you can define the minimum and maximum values that should be accepted. If you designate some kind of ID to your sequences, you can use the bounded number to define that it should be at least 1 and max 99999 if that is the range of your IDs.
- **Decimal number.** Same as number, but it will also accept decimal numbers.
- **Bounded decimal number.** Same as bounded number, but it will also accept decimal numbers.

When you click **OK**, the attribute will appear in the list to the left. Clicking the attribute will allow you to see information on its type in the panel to the right.

Lists are a little special, since you have to define the items in the list. When you choose to add the list attribute in the left side of the dialog, you can define the items of the list in the panel to the right by clicking **Add Item** (+) (see figure 8.3).

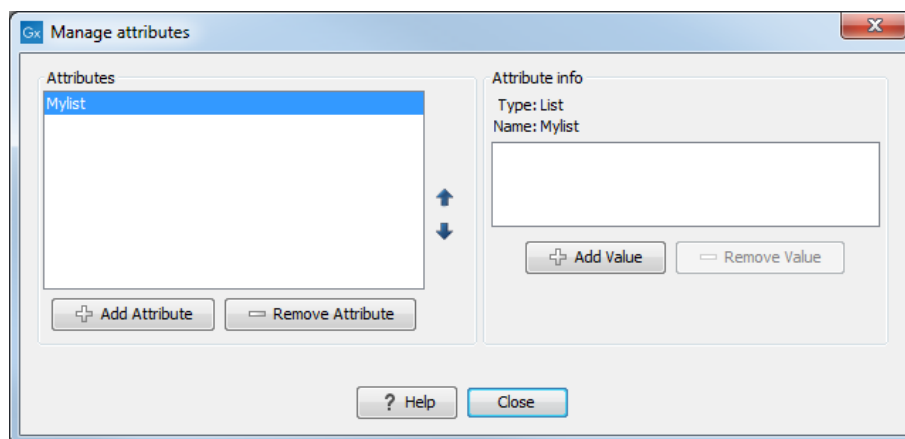


Figure 8.3: Defining items in a list.

Remove items in the list by pressing **Remove Item** (−).

Removing attributes To remove an attribute, select the attribute in the list and click **Remove Attribute** (≡). This can be done without any further implications if the attribute has just been created, but if you remove an attribute where values have already been given for elements in the data location, it will have implications for these elements: The values will not be removed, but they will become static, which means that they cannot be edited anymore.

If you accidentally removed an attribute and wish to restore it, this can be done by creating a new attribute of exactly the same name and type as the one you removed. All the "static" values will now become editable again.

When you remove an attribute, it will no longer be possible to search for it, even if there is "static" information on elements in the data location.

Renaming and changing the type of an attribute is not possible - you will have to create a new one.

Changing the order of the attributes You can change the order of the attributes by selecting an attribute and click the **Up** and **Down** arrows in the dialog. This will affect the way the attributes are presented for the user.

8.1.1 Filling in values

When a set of attributes has been created (as shown in figure 8.4), the end users can start filling in information.

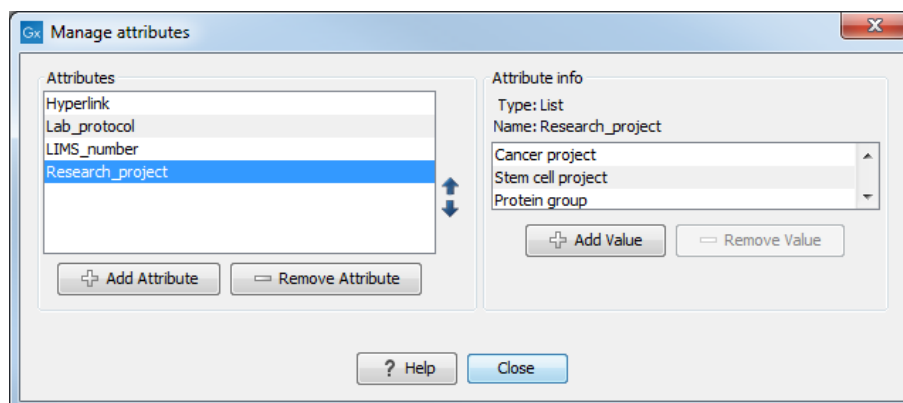


Figure 8.4: A set of attributes defined in the attribute manager.

This is done in the element info view:

right-click a sequence or another element in the Navigation Area | Show (☞) | Element info (📄)

This will open a view similar to the one shown in figure 8.5.

You can now enter the appropriate information and **Save**. When you have saved the information, you will be able to search for it (see below).

Note that the element (e.g. sequence) needs to be saved in the data location before you can edit the attribute values.

When nobody has entered information, the attribute will have a "Not set" written in red next to the attribute (see figure 8.6).

1WCL_A x

Fixed Fields

- ▼ Name Edit 1WCL_A
- ▼ Description Edit
- ▼ Metadata Refresh This element is not associated with any metadata
- Local Attribute Fields
- ▼ Research_project Clear Cancer project
- ▼ Hyper_link Edit Clear <http://www.clcbio.com>
- ▼ Is_confirmed Clear ☒
- ▼ Lab_protocol Edit Clear P123
- ▼ LIMS_number Clear 412
- ▼ Location Clear Lab 23
- ▼ Patient_number Clear 651

Figure 8.5: Adding values to the attributes.

Local Attribute Fields

- ▼ Research_project Clear Cancer project
- ▼ Hyper_link Edit Clear Not set

Figure 8.6: An attribute which has not been set.

This is particularly useful for attribute types like checkboxes and lists where you cannot tell, from the displayed value, if it has been set or not. Note that when an attribute has not been set, you cannot search for it, even if it looks like it has a value. In figure 8.6, you will *not* be able to find this sequence if you search for research projects with the value "Cancer project", because it has not been set. To set it, simply click in the list and you will see the red "Not set" disappear.

If you wish to reset the information that has been entered for an attribute, press "Clear" (written in blue next to the attribute). This will return it to the "Not set" state.

The **Folder editor**, invoked by pressing **Show** on a given folder from the context menu, provides a quick way of changing the attributes of many elements in one go (see the Workbench manuals at <http://www.qiagenbioinformatics.com/support/manuals/>).

8.1.2 What happens when a clc object is copied to another data location?

The user supplied information, which has been entered in the **Element info**, is attached to the attributes that have been defined in this particular data location. If you copy the sequence to another data location or to a data location containing another attribute set, the information will become fixed, meaning that it is no longer editable and cannot be searched for. Note that attributes that were "Not set" will disappear when you copy data to another location.

If the element (e.g. sequence) is moved back to the original data location, the information will again be editable and searchable.

If the e.g. Molecule Project or Molecule Table is moved back to the original data location, the information will again be editable and searchable.

8.1.3 Searching

When an attribute has been created, it will automatically be available for searching. This means that in the **Local Search** (🔍), you can select the attribute in the list of search criteria (see figure 8.7).

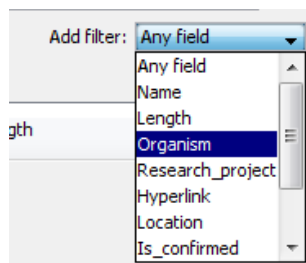


Figure 8.7: The attributes from figure 8.4 are now listed in the search filter.

It will also be available in the **Quick Search** below the **Navigation Area** (press Shift+F1 (Fn+Shift+F1 on Mac) and it will be listed - see figure 8.8).

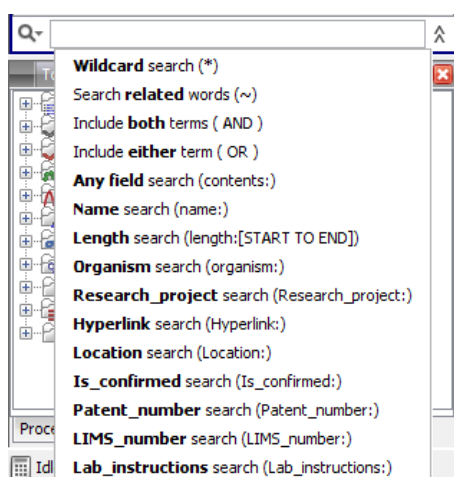


Figure 8.8: The attributes from figure 8.4 are now available in the Quick Search as well.

Read more about search in one of the Workbench manuals: http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Local_search.html.

Chapter 9

Appendix

9.1 Troubleshooting

9.1.1 Check setup

To check your server is set up correctly, run the **check setup** tool. To do this:

- Log in on the web interface of the server as an administrator.
- Click the **check setup** link in the top right hand corner.
- Click on the **Generate Diagnostics Report** button in the window that appears.

When the report is ready, a list of tests performed is provided, as shown in figure 9.1.

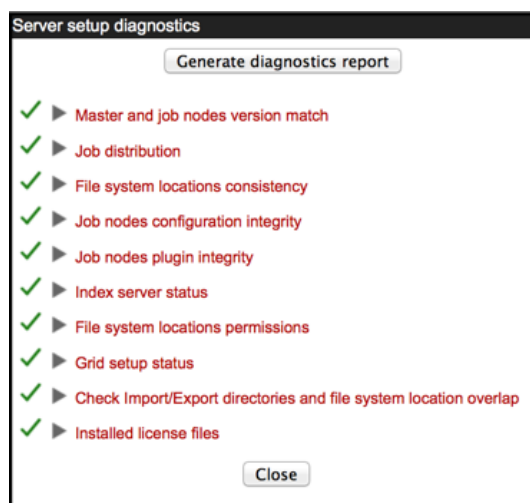


Figure 9.1: Check system. Tests that passed are marked with a green check mark. Tests that failed are marked with a red X.

Tests that passed are marked with a green check mark. Tests that failed are marked with a red X. Click on any of the tests listed to see more information about the test.

Additional notes:

- A green check mark is presented beside "List license files" when the contents of the "licenses" folder in the installation area of the *CLC Bioinformatics Database* could be listed. Click on this item to see a list of the licenses found. The products and versions supported by the licenses found are reported. Information about expired licenses is also presented. See figure 9.2.
- A green check mark is presented beside "Grid setup status" in two cases:
 - You have configured a grid setup and it is configured correctly.
 - You have not configured a grid setup.



Figure 9.2: Click on the *List license files* item in the report to see the list of the license files found in the licenses subfolder of the installation area. Products and versions supported are reported, and any expired license is noted with red text.

9.1.2 Bug reporting

If there are problems regarding the installation and configuration of the server, please contact ts-bioinformatics@qiagen.com.

When contacting ts-bioinformatics@qiagen.com regarding problems on the *CLC Bioinformatics Database*, you will often be asked for additional information about the server setup. You can easily send the necessary information by submitting a bug report:

Log in to the web interface of the server as administrator | report a bug (at the top right corner) | Enter relevant information with as much detail as possible | Submit Bug Report

You can see the bug report dialog in 9.3.

The bug report includes the following information:

- Log files
- A subset of the audit log showing the last events that happened on the server
- Configuration files of the server configuration

In a job node setup you can include the information from the job nodes by checking the **Include comprehensive job node info** checkbox in the **Advanced** part of the dialog.

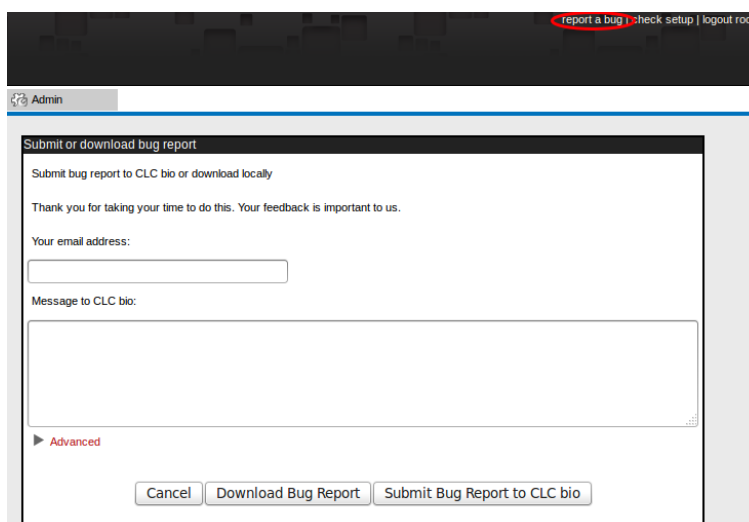
The screenshot shows a web interface for submitting or downloading a bug report. At the top, there is a navigation bar with links: 'report a bug' (highlighted with a red circle), 'check setup', and 'logout root'. Below this is a header with 'Admin' and a gear icon. The main content area is titled 'Submit or download bug report'. It contains a sub-header 'Submit bug report to CLC bio or download locally', a thank-you message 'Thank you for taking your time to do this. Your feedback is important to us.', and a form with two fields: 'Your email address:' and 'Message to CLC bio:'. Below the form is a red 'Advanced' link. At the bottom, there are three buttons: 'Cancel', 'Download Bug Report', and 'Submit Bug Report to CLC bio'.

Figure 9.3: Submitting a bug report.

If the server has access to the internet, you can **Submit Bug Report** to send the report to QIAGEN Bioinformatics Support. If the server does not have access to the internet, click on **Download bug report** to create a zip file containing the same information that you can attach to an email you send to ts-bioinformatics@qiagen.com from a machine connected to the network.

Note that the process of gathering the information for the bug report can take a while, especially for job node setups. If a Workbench user experiences a server-related error, it is also possible to submit a bug report from a Workbench error dialog if they are presented with one. The same archive is included as when submitting a bug report from the server web interface.

No password information is included in the bug report.

All data sent to ts-bioinformatics@qiagen.com is treated confidentially.

9.2 Database configurations

To use an SQL database for data management with the *CLC Bioinformatics Database*, the appropriate JDBC driver for your database system must be installed. This is described in section [9.2.1](#).

For MySQL installations, please also refer to section [9.2.2](#) for information specific to MySQL database configuration.

9.2.1 Installing JDBC drivers

The general steps for installing JDBC drivers for use with CLC software are:

1. Download the appropriate JDBC driver for your database system from the provider. Further details about this are provided below.
2. Place the driver into the `userlib` directory in the installation area of the *CLC Bioinformatics Database*.
 - For a CLC job node setup, the JDBC driver file must be placed into the `userlib` folder

under the *CLC Bioinformatics Database* installation area on **the master node and also on each of the job nodes**.

- For a grid node setup, the driver file only needs to be placed into the `userlib` folder of the master *CLC Bioinformatics Database* installation area.

3. Restart the *CLC Bioinformatics Database* software.

- For a job node setup, restart the master server and each of the job nodes.
- For a grid setup, only the master CLC Server needs to be restarted. This will cause the changes to be deployed to the grid workers.

Getting JDBC drivers

Information on obtaining JDBC drivers for supported DBMS systems is provided in this section. After obtaining the driver, the general instructions given above should be followed to complete the driver installation.

MySQL JDBC Drivers

1. Go to <http://dev.mysql.com/downloads/connector/j/> to download the driver.
2. Choose the option **Platform Independent** when selecting a platform.
3. After clicking on the button to Download, you can login if you already have an Oracle Web account, or you can just click on the link that says `No thanks, just start my download` further down the page.
4. Uncompress the downloaded file and move the driver file, which will have a name of this form: `mysql-connector-java-X.X.XX-bin.jar`, to the folder called `userlib`, found in the installation area of the *CLC Bioinformatics Database*.

PostgreSQL JDBC Drivers

1. Go to <https://jdbc.postgresql.org/download.html> and download the relevant driver.
2. Place it in the folder called `userlib`, found in the installation area of the *CLC Bioinformatics Database*.

Microsoft SQL Server

1. Go to <https://docs.microsoft.com/en-us/sql/connect/jdbc/download-microsoft-jdbc-driver-for-sql-server> and download the relevant driver.
2. Follow the installation instructions given for the driver to the point where the downloaded file has been unpacked.
3. Move the driver jar file compatible with java 1.8 to the folder called `userlib`, found in the installation area of the *CLC Bioinformatics Database*.

Oracle JDBC Drivers

1. Go to <http://www.oracle.com/technetwork/database/application-development/jdbc/downloads/index.html>.
2. Select the relevant version for your Oracle database version.
You will need an Oracle account to download the driver.
3. Move the driver jar file to the folder called `userlib`, found in the installation area of the *CLC Bioinformatics Database*.

9.2.2 Configuring MySQL

We recommend basing your MySQL configuration on the example configuration file `my-large.cnf` included in the MySQL distribution.

In addition the following changes should be made:

- Increase the value of the `max_allowed_packet` setting to support transfer of large binary objects to and from the database will be supported. We recommend this setting:
`max_allowed_packet = 64M`
- Ensure InnoDB is available and configured. This is necessary for the MySQL instance to work properly with the *CLC Bioinformatics Database*.
- Enable the options in the InnoDB section of the configuration as outlined below:

```
# You can set .._buffer_pool_size up to 50 - 80 %
# of RAM but beware of setting memory usage too high
innodb_buffer_pool_size = 256M
innodb_additional_mem_pool_size = 20M
# Set .._log_file_size to 25 % of buffer pool size
innodb_log_file_size = 64M
innodb_log_buffer_size = 8M
innodb_flush_log_at_trx_commit = 1
innodb_lock_wait_timeout = 50
```

Additionally, there appears to be a bug in certain versions of MySQL that can cause the cleanup of the query cache to take a very long time. If you experience this, please disable the query log by setting the `query_cache_size` option to 0: `query_cache_size= 0`

9.3 SSL and encryption

The *CLC Bioinformatics Database* supports SSL communication between the Server and its clients (i.e. Workbenches or the *CLC Server Command Line Tools*). This is particularly relevant if the server is accessible over the internet as well as on a local network.

The default configuration of the server does not use SSL.

9.3.1 Enabling SSL on the server

A **server certificate** is required before SSL can be enabled on the *CLC Bioinformatics Database*. This is usually obtained from a *Certificate Authority* (CA) like Thawte or Verisign (see http://en.wikipedia.org/wiki/Certificate_authorities).

A **signed certificate** in a `pkcs12` keystore file is also needed. The keystore file is either provided by the CA or it can be generated from the private key used to request the certificate and the signed-certificate file from the CA (see section 9.3.1).

Copy the keystore file to the `conf` subdirectory of the *CLC Bioinformatics Database* installation folder.

Next, the `server.xml` file in the `conf` subdirectory of the *CLC Bioinformatics Database* installation folder has to be edited to enable SSL-connections. Add text like the following text to the `server.xml` file:

```
<Connector port="8443" protocol="HTTP/1.1" SSLEnabled="true"
    maxThreads="150" scheme="https" secure="true"
    clientAuth="false" sslProtocol="TLS"
    keystoreFile="conf/keystore.pkcs12" keystorePass="tomcat"
    keystoreType="PKCS12"
/>
```

Replace `keystore.pkcs12` with the name of your keystore file, and replace `tomcat` with the password for your keystore.

The above settings make SSL available on port 8443. The standard (non-SSL) port would still be 7777, or whatever port number you have configured it to.

Self-signed certificates can be generated if only connection encryption is needed. See http://www.akadia.com/services/ssh_test_certificate.html for further details.

Creating a PKCS12 keystore file

If the certificate is not supplied in a `pkcs12` keystore file, it can be put into one by combining the private key and the signed certificate obtained from the CA by using `openssl`:

```
openssl pkcs12 -export -out keystore.pkcs12 -inkey private.key -in certificate.crt -name "tomcat"
```

This will take the private key from the file `private.key` and the signed certificate from `certificate.crt` and generate a `pkcs12`-store in the `keystore.pkcs12` file.

9.3.2 Logging in using SSL from the Workbench

When the Workbench connects to the *CLC Bioinformatics Database* it automatically detects if Secure Socket Layer (SSL) should be used on the port it is connecting to or not.

If SSL is detected, the server's certificate will be verified and a warning is displayed if the certificate is not signed by a recognized Certificate Authority (CA) as shown in figure 9.4.

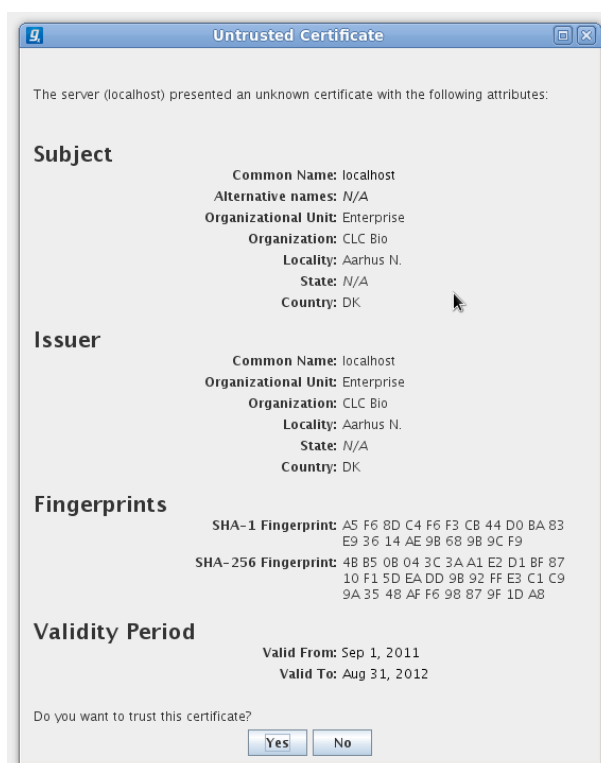


Figure 9.4: A warning is shown when the certificate is not signed by a recognized CA.

When such an "unknown" certificate has been accepted once, the warning will not appear again. It is necessary to log in again once the certificate has been accepted.

When logged into a server, information about the connection can be viewed by hovering the connection icon on the status-panel as shown in figure 9.5.

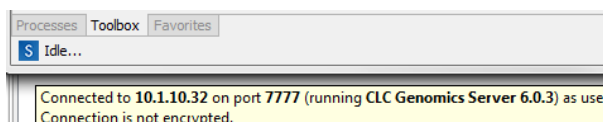


Figure 9.5: Showing details on the server connection by placing the mouse on the globe.

The icon is gray when the user is not logged in, and a pad lock is overlaid when the connection is encrypted via SSL.

9.3.3 Logging in using SSL from the CLC Server Command Line Tools

The CLC Server Command Line Tools will also automatically detect and use SSL if present on the port it connects to. If the certificate is untrusted the `clcserver` program will refuse to login:

```
./clcserver -S localhost -U root -W default -P 8443
Message: Trying to log into server
Error: SSL Handshake failed. Check certificate.
Option          Description
-----
-A <Command>    Command to run. If not specified the list of commands on the server will be returned.
-C <Integer>    Specify column width of help output.
-D <Boolean>    Enable debug mode (default: false)
-G <Grid Preset value> Specify to execute on grid.
-H             Display general help.
-I <Algorithm Command> Get information about an algorithm
-O <File>       Output file.
-P <Integer>    Server port number. (default: 7777)
-Q <Boolean>    Quiet mode. No progress output. (default: false)
```

```

-S <String>          Server hostname or IP-address of the CLC Server.
-U <String>          Valid username for logging on to the CLC Server
-V                  Display version.
-W <String>          Clear text password or domain specific password token.

```

In order to trust the certificate the `clcsserverssslstore` tool must be used:

```

./clcsserverssslstore -S localhost -U root -W default -P 8443
The server (localhost) presented an untrusted certificate with the following attributes:
SUBJECT
=====
Common Name           : localhost
Alternative Names     : N/A
Organizational Unit   : Enterprise
Organization          : CLC Bio
Locality              : Aarhus N.
State                 : N/A
Country               : DK

ISSUER
=====
Common Name           : localhost
Organizational Unit   : Enterprise
Organization          : CLC Bio
Locality              : Aarhus N.
State                 : N/A
Country               : DK

FINGERPRINTS
=====
SHA-1                 : A5 F6 8D C4 F6 F3 CB 44 D0 BA 83 E9 36 14 AE 9B 68 9B 9C F9
SHA-256               : 4B B5 0B 04 3C 3A A1 E2 D1 BF 87 10 F1 5D EA DD 9B 92 FF E3 C1 C9 9A 35 48 AF F6 98 87 9F 1D A8

VALIDITY PERIOD
=====
Valid From            : Sep 1, 2011
Valid To               : Aug 31, 2012
Trust this certificate? [yn]

```

Once the certificate has been accepted, the `clcserver` program is allowed to connect to the server.

9.4 Third party libraries

The *CLC Bioinformatics Database* includes a number of third party libraries.

Please consult the files named NOTICE and LICENSE in the server installation directory for the legal notices and acknowledgements of use.

For the code found in this product that is subject to the Lesser General Public License (LGPL) you can receive a copy of the corresponding source code by sending a request to our support team at ts-bioinformatics@qiagen.com.

Bibliography