Building from Source Code

Table of Contents:

- Prerequisites
  - Getting and Compiling Dependencies
- Clone and Build OpenCGA
  - Building with Maven

Prerequisites

In order to build OpenCGA from source code you must first get the source code of OpenCGA from GitHub, most of the dependencies - including OpenCB dependencies - will be fetched from Maven Central Repository, however in some scenarios OpenCB dependencies will need to be built from GitHub source code. Compiling and building processes are carried out by **Apache Maven**. The following tools are required for successful build:

- Java 1.8.0_60+
- Apache Maven

You can learn how to install them in this section at **Installation Guide > Server Configuration**.

Getting and Compiling Dependencies

OpenCGA as any other software has dependencies, some of them come from other OpenCB projects such as CellBase while others are third-party dependencies such as MongoDB. All OpenCGA ***stable*** releases are always merged and tagged at ***master*** branch (users are encouraged to use latest stable release for production), you can find all releases at OpenCGA Releases. We guarantee that all the dependencies needed for building stable releases are deployed at Maven Central Repository, this is true for both OpenCB and third-party dependencies. Therefore for building a stable release you only need to clone OpenCGA repository itself since all the dependencies will be fetched form Maven Central repository.

This is different for development branches. Active OpenCGA development is carried out at ***develop*** branc h, in this branch third-party dependencies will be still fetched from Maven Central Repository **but** this is not true for OpenCB dependencies since it is very likely they are still in development and therefore they are not deployed. Keep in mind that we only guarantee that *develop* compiles and that bugs are expected, use this branch for development or for testing new functionalities. So, for building ***develop*** bran ch you may require to download and install the following OpenCB repositories in this order:

- *java-common-libs:*https://github.com/opencb/java-common-libs (branch 'develop')
- *biodata*: https://github.com/opencb/biodata (branch 'develop')
- *cellbase*: https://github.com/opencb/cellbase (branch 'develop')
- *hpg-bigdata*: https://github.com/opencb/hpg-bigdata (branch 'develop')

As you can see one of our rules is that *develop* branch of all major applications such as OpenCGA and CellBase always depend on *develop* branches. So, if you really want to build *develop* the you can clone and build dependencies by executing:

Clone Dependencies

```
## Clone develop branc
git clone -b develop https://github.com/opencb/java-common-libs.git
git clone -b develop https://github.com/opencb/biodata.git
git clone -b develop https://github.com/opencb/cellbase.git
git clone -b develop https://github.com/opencb/hpg-bigdata.git

## Now you can execute the following command in each of the folders the
specified order above
mvn clean install -DskipTests
```

Clone and Build OpenCGA

You can clone OpenCGA from GitHub by executing:

Shell

```
## Latest stable version
git clone -b master https://github.com/opencb/opencga.git

## Develop branch, for this to work remember to clone and build OpenCB
dependencies (see above)
git clone -b develop https://github.com/opencb/opencga.git
```

Building with Maven

OpenCGA allows to customise many variables in the configuration files, in order to make easier the building and configuration of OpenCGA we rely on Maven Properties that can be defined in file `~/.m2/settings.xml`. During the building all these properties will be injected automatically in the configuration files so users do not have to manually change all configuration values. **Note** this is only possible when building OpenCGA from source code, if you download the binary version you will have to manually set up all configuration variables.

An example of that file can be found in the README and below. The description of each property can be found below:

- **OPENCGA.CATALOG.DB.HOSTS**: This property should be configured with the host and port of the MongoDB installation. By default, for development purposes, we have it set with "localhost:27017".
- **OPENCGA.CATALOG.DB.USER**: This property should only be set if the MongoDB needs authentication. In this case, this property will contain the user name with permissions for the database. *This can be left empty in any case. The admin will be able to set this credentials using the command line.
- **OPENCGA.CATALOG.DB.PASSWORD**: This property should only be set if the MongoDB needs authentication. In this case, this property will contain the password of the user with permissions for the database. *Like in the user property, this can be left empty. The admin will be able to set this credentials using the command line.
- **OPENCGA.INSTALLATION.DIR**: This property is extremely important when using Tomcat to deploy the webservices. This property will have to point to the final OpenCGA installation directory after everything has been built. This property will be used by Tomcat to locate the configuration files. If this is not properly set, none of the webservices will work. Default: /opt/opencga.
- **OPENCGA.USER.WORKSPACE**: In Catalog, users are allowed to build their own directory structure, upload their own files, run analysis, etc. This path should be pointing to a physical location where Catalog will be storing those files and directory structure. By default, we normally put it in a folder called "sessions" within the installation directory (file:///opt/opencga/sessions/).
- **OPENCGA.STORAGE.VARIANT.DB.HOSTS**:
- **OPENCGA.STORAGE.VARIANT.DB.USER**:
- **OPENCGA.STORAGE.VARIANT.DB.PASSWORD**:
- **OPENCGA.ANALYSIS.EXECUTION.MANAGER**: OpenCGA Catalog allows users to run jobs. This property indicates how the jobs will be launched. At the moment we only support two types: LOCAL to run the jobs locally in a thread or SGE to run the jobs using Sun Grid Engine. More queuing systems will be supported soon.
- **OPENCGA.CLIENT.HOST**: This property should be pointing to the URL where the webservices will be available. For development purposes, the default is http://localhost:8080/opencga/. This property is read by the command line opencga.sh in order to communicate with the webservices.
- **OPENCGA.CELLBASE.REST.HOST**: URL to be used for Variant Annotation.

You can copy this example to *~/.m2/settings.xml*:

Update

The next XML code has been updated on 15 Jan 2017, please make sure you update your *settings.xml*, sorry for any inconvenience caused.

```
    <?xml version="1.0" encoding="UTF-8"?>
    <settings xmlns="http://maven.apache.org/SETTINGS/1.0.0"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://maven.apache.org/SETTINGS/1.0.0
http://maven.apache.org/xsd/settings-1.0.0.xsd">
        <profiles>
            <profile>
                <id>custom-config</id>
                <activation>
                    <activeByDefault>true</activeByDefault>
                </activation>
                <properties>
                    <opencga.war.name>opencga-${opencga.version}</opencga.
war.name>

                    <!-- General -->
                    <OPENCGA.INSTALLATION.DIR>/opt/opencga</OPENCGA.
INSTALLATION.DIR>
                    <OPENCGA.USER.WORKSPACE>file:///opt/opencga/sessions/<
/OPENCGA.USER.WORKSPACE>
                    <OPENCGA.JOBS.DIR>${OPENCGA.USER.WORKSPACE}/jobs/<
```

```
/OPENCGA.JOBS.DIR>
                    <OPENCGA.DB.PREFIX>opencga</OPENCGA.DB.PREFIX>
                    <OPENCGA.EXECUTION.MODE>LOCAL</OPENCGA.EXECUTION.MODE>

                    <!-- Client -->
                    <OPENCGA.CLIENT.REST.HOST>http://localhost:8080
/${opencga.war.name}</OPENCGA.CLIENT.REST.HOST>
                    <OPENCGA.CLIENT.GRPC.HOST>http://localhost:9091<
/OPENCGA.CLIENT.GRPC.HOST>
                    <OPENCGA.CLIENT.ORGANISM.SCIENTIFIC_NAME>Homo sapiens<
/OPENCGA.CLIENT.ORGANISM.SCIENTIFIC_NAME>
                    <OPENCGA.CLIENT.ORGANISM.COMMON_NAME>human</OPENCGA.
CLIENT.ORGANISM.COMMON_NAME>
                    <OPENCGA.CLIENT.ORGANISM.TAXONOMY_CODE>9606</OPENCGA.
CLIENT.ORGANISM.TAXONOMY_CODE>
                    <OPENCGA.CLIENT.ORGANISM.ASSEMBLY></OPENCGA.CLIENT.
ORGANISM.ASSEMBLY>

                    <OPENCGA.SERVER.REST.PORT>9090</OPENCGA.SERVER.REST.
PORT>
                    <OPENCGA.SERVER.GRPC.PORT>9091</OPENCGA.SERVER.GRPC.
PORT>
                    <OPENCGA.MONITOR.PORT>9092</OPENCGA.MONITOR.PORT>

                    <!-- Catalog -->
                    <OPENCGA.CATALOG.DB.HOSTS>localhost:27017</OPENCGA.
CATALOG.DB.HOSTS>
                    <OPENCGA.CATALOG.DB.USER></OPENCGA.CATALOG.DB.USER>
                    <OPENCGA.CATALOG.DB.PASSWORD></OPENCGA.CATALOG.DB.
PASSWORD>
                    <OPENCGA.CATALOG.DB.AUTHENTICATION_DATABASE></OPENCGA.
CATALOG.DB.AUTHENTICATION_DATABASE>
                    <OPENCGA.CATALOG.DB.CONNECTIONS_PER_HOST>20</OPENCGA.
CATALOG.DB.CONNECTIONS_PER_HOST>

                    <!-- Storage -->
                    <OPENCGA.STORAGE.DEFAULT_ENGINE>mongodb</OPENCGA.
STORAGE.DEFAULT_ENGINE>
                    <OPENCGA.STORAGE.CACHE.HOST>localhost:6379</OPENCGA.
STORAGE.CACHE.HOST>
                    <OPENCGA.STORAGE.SEARCH.HOST>http://localhost:8983/solr
/</OPENCGA.STORAGE.SEARCH.HOST>
                    <OPENCGA.STORAGE.STUDY_METADATA_MANAGER></OPENCGA.
STORAGE.STUDY_METADATA_MANAGER>

                    <!-- Storage Variants general -->
                    <OPENCGA.STORAGE.VARIANT.DB.HOSTS>localhost:27017<
/OPENCGA.STORAGE.VARIANT.DB.HOSTS>
                    <OPENCGA.STORAGE.VARIANT.DB.USER></OPENCGA.STORAGE.
VARIANT.DB.USER>
                    <OPENCGA.STORAGE.VARIANT.DB.PASSWORD></OPENCGA.STORAGE.
VARIANT.DB.PASSWORD>

                    <!-- Storage Alignments general -->
                    <OPENCGA.STORAGE.ALIGNMENT.DB.HOSTS>localhost:27017<
/OPENCGA.STORAGE.ALIGNMENT.DB.HOSTS>
                    <OPENCGA.STORAGE.ALIGNMENT.DB.USER></OPENCGA.STORAGE.
ALIGNMENT.DB.USER>
                    <OPENCGA.STORAGE.ALIGNMENT.DB.PASSWORD></OPENCGA.
STORAGE.ALIGNMENT.DB.PASSWORD>

                    <!-- Storage-mongodb -->
                    <OPENCGA.STORAGE.MONGODB.VARIANT.DB.
AUTHENTICATION_DATABASE></OPENCGA.STORAGE.MONGODB.VARIANT.DB.
AUTHENTICATION_DATABASE>
                    <OPENCGA.STORAGE.MONGODB.VARIANT.DB.
CONNECTIONS_PER_HOST>20</OPENCGA.STORAGE.MONGODB.VARIANT.DB.
CONNECTIONS_PER_HOST>

                    <!-- Storage-hadoop -->
                    <!--If empty, will use the ZOOKEEPER_QUORUM read from
```

```
the hbase configuration files-->
                    <OPENCGA.STORAGE.HADOOP.VARIANT.DB.HOSTS></OPENCGA.
STORAGE.HADOOP.VARIANT.DB.HOSTS>
                    <OPENCGA.STORAGE.HADOOP.VARIANT.DB.USER></OPENCGA.
STORAGE.HADOOP.VARIANT.DB.USER>
                    <OPENCGA.STORAGE.HADOOP.VARIANT.DB.PASSWORD></OPENCGA.
STORAGE.HADOOP.VARIANT.DB.PASSWORD>
                    <OPENCGA.STORAGE.HADOOP.VARIANT.HBASE.NAMESPACE><
/OPENCGA.STORAGE.HADOOP.VARIANT.HBASE.NAMESPACE>
                    <OPENCGA.STORAGE.HADOOP.VARIANT.ARCHIVE.TABLE.
PREFIX>${OPENCGA.DB.PREFIX}_study</OPENCGA.STORAGE.HADOOP.VARIANT.ARCHIVE.
TABLE.PREFIX>

                    <!-- Email server -->
                    <OPENCGA.MAIL.HOST></OPENCGA.MAIL.HOST>
                    <OPENCGA.MAIL.PORT></OPENCGA.MAIL.PORT>
                    <OPENCGA.MAIL.USER></OPENCGA.MAIL.USER>
                    <OPENCGA.MAIL.PASSWORD></OPENCGA.MAIL.PASSWORD>

                    <!-- cellbase -->
                    <OPENCGA.CELLBASE.REST.HOST>http://bioinfo.hpc.cam.ac.
uk/cellbase/</OPENCGA.CELLBASE.REST.HOST>
                                        <OPENCGA.CELLBASE.VERSION>v4<
/OPENCGA.CELLBASE.VERSION>
                    <OPENCGA.CELLBASE.DB.HOST>localhost:27017</OPENCGA.
CELLBASE.DB.HOST>
                    <OPENCGA.CELLBASE.DB.USER></OPENCGA.CELLBASE.DB.USER>
                    <OPENCGA.CELLBASE.DB.PASSWORD></OPENCGA.CELLBASE.DB.
PASSWORD>
                    <OPENCGA.CELLBASE.DB.AUTHENTICATION_DATABASE></OPENCGA.
CELLBASE.DB.AUTHENTICATION_DATABASE>
                    <OPENCGA.CELLBASE.DB.
READ_PREFERENCE>secondaryPreferred</OPENCGA.CELLBASE.DB.READ_PREFERENCE>
                </properties>
            </profile>
        </profiles>
    </settings>
```

After creating this and configuring a default profile, you can build OpenCGA by executing the following command from the root of the cloned repository:

```
$ mvn clean install -DskipTests -P custom-config
```

The first time this command can take some minutes since it has to fetched and store locally all the dependencies, next builds will be much faster. After successful building, user should find the following file structure under a build folder:

Shell

```
build
 bin
    obsolete
    opencga-admin.sh
    opencga-analysis.sh
    opencga-env.sh
    opencga.sh
 conf
    client-configuration.yml
    configuration.yml
    log4j.properties
    storage-configuration.yml
 examples
    1k.chr1.phase3_shapeit2_mvncall_integrated_v5.20130502.genotypes.vcf.gz
    20130606_g1k.ped
    opencga-index.sh
    opencga-storage-fetch.sh
 libs
    activation-1.1.jar
    .....
    zookeeper-3.4.6.jar
 LICENSE
 opencga-1.0.0-rc4.war
 README.md
 test
    bin
    dependencies
    fitnesse
    README.md
 tools
     affy-expression-normalization
     .....
     variant

42 directories, 181 files
```