Installation

You do not need to install CellBase to run queries. See Using CellBase for more information on how to search using CellBase.

If you do need a local installation of CellBase, please make sure that the server(s) have all dependencies installed then you can configure and complete the CellBase build.

Table of Contents:

- Overview
- Step 1 - Configuring the Server
  - Hardware
  - Software Dependencies
- Step 2 - Getting the CellBase code
  - Building from Sources
  - Download Binaries
  - Docker
- Step 3 - Downloading the data
- Step 4 - Building the data
- Step 5 - Loading the data
- Step 6 - Web services
  - REST API
  - Tomcat

Overview

Building a CellBase instance has three stages:

Stage	Description
Download **	Downloads the data files for the specified data sets
Build **	Parses the downloaded data files, generates JSON objects, e.g. gene.json
Load	Loads the generated JSON objects into the Mongo database

First, you will **download** a set of raw files from several data sources. These raw files shall contain the core data that will populate the Cellbase knowledgebase. Then, you will **build** the JSON documents that should be **loaded** into the Cellbase knowledgebase. These three stages are described in detail below.

** We have already downloaded and processed these data, and the resulting JSON documents are available through our FTP server. For those users who wish to skip these two sections, directly download json documents from http://bioinfo.hpc.cam.ac.uk/downloads/cellbase/v4/homo_sapiens_grch37/mongodb/ and jump to the Load Data section.

Step 1 - Configuring the Server

Before you can start building CellBase, you must first install all required software dependencies.

Hardware

Which sort of hardware you need depends on how much data you need, query load, etc. A full CellBase instance is 1 TB of data, but loading only genomic data is XXX GB. Also loading and querying data is very resource intensive, we recommend at least 8 GB of RAM.

Software Dependencies

Below are the software dependencies required by CellBase.

Software	Version	Purpose
Java	8	Build and use CellBase
MongoDB	3.6	Database
Tomcat	8.5x	REST API
Docker	18	Building Ensembl

- Java - we recommend you use the OpenJDK.
- MongoDB - put your mongo credentials in settings.xml ???
- Tomcat - put your tomcat credentials in settings.xml ???
- Docker - CellBase uses docker to manager the Perl modules required to query Ensembl's Perl API.

Step 2 - Getting the CellBase code

There are three main ways to get CellBase for installation:

- You can download the source code from GitHub and use **Apache Maven** to compile and build it.
- Or you can download a *prebuilt* binary from the **CellBase GitHub Releases** web page, notice that only *stable* and *pre-releases* are tagged and prebuilt.
- Or you can use our Docker image containing the CellBase binaries.

Here you can learn more about these options.

Building from Sources

Although most users will use stable *prebuilt* binaries (see below) there is still the need for different users to compile and build CellBase, for instance to test a development version. You can learn how to build from the source code at **Installation Guide > Building from Source Code**.

Download Binaries

You can download *stable* and *pre-release* (beta and release candidate) versions from **CellBase GitHub Releases**. You will find a *tar.gz* file with the name of *cellbase* and the *version,* for instance to download CellBase 4.7.1 you can go to the GitHub Release at:

https://github.com/opencb/cellbase/releases/tag/v4.7.1

Download the file *cellbase-4.7.1.tar.gz* from the *Downloads* section.

Docker

We also have docker images containing the CellBase binaries, see CellBase at DockerHub. This docker image expects a running MongoDB instance.

Step 3 - Downloading the data

Run this command to download all the data:

```
$ ./build/bin/cellbase-admin.sh download -d gene -s hsapiens
```

See Download Sources for the details on all the data that's available to download.

Step 4 - Building the data

Run this command to download all the data:

```
$ ./build/bin/cellbase-admin.sh build -d gene -s hsapiens
```

See Building the CellBase database for the details on how to build.

Step 5 - Loading the data

Run this command to download all the data:

```
$ ./build/bin/cellbase-admin.sh load -d gene -s hsapiens
```

See Load Data for the details on all the data that's available to download.

Step 6 - Web services

CellBase has two options:

REST API

```
$ ./build/bin/cellbase-admin.sh server --start
```

View the API here: http://localhost:9090/cellbase-5.0.0-SNAPSHOT/webservices/

Tomcat

Make sure Tomcat is running, then copy the WAR file the build has created for you into Tomcat's webapps directory:

cp ./build/cellbase-5.0.0-SNAPSHOT.war ~/apache-tomcat-8.5.47/webapps/

You should see the generated API docs here:

http://localhost:8080/cellbase-5.0.0-SNAPSHOT/webservices/