

Load Data

After you have generated the JSON data files (see [Building the CellBase database](#)), you can now load them in your Mongo database.

Configure Database Credentials

Before loading the data into the database, your settings.xml file must have been appropriately configured indicating the database host names, ports, user and password. Here are the expected parameters:

name	default
CELLBASE.DB.MONGODB.HOST	localhost:27017
CELLBASE.DB.USER	<empty>
CELLBASE.DB.PASSWORD	<empty>
CELLBASE.SERVER.REST.PORT	9090

Table of Contents:

- [Configure Database Credentials](#)
- [Load JSON files into MongoDB](#)

Load JSON files into MongoDB

Use the CellBase CLI to load the data models. For example, to load all human (GRCh37) data models from the /tmp/data/cellbase/v4/homo_sapiens_grch37/mongodb/ created in section "Build Data Models", into the cellbase_hsapiens_grch37_v4 database:

```
cellbase/build/bin$ ./cellbase.sh load -d variation --database
cellbase_hsapiens_grch37_v4 -i /mnt/data/downloads/cellbase/v4
/homo_sapiens_grch37/mongodb/ -L debug -Dmongodb-index-folder=/home
/caftero/appl/dev/cellbase/cellbase-app/app/mongodb-scripts/
```

Please, note that the whole loading and indexing process may need ~24h to complete, depending on the available hardware.

After successful load of all data, the corresponding database shall look like:

```
$ mongo mongodb-dev/cellbase_hsapiens_grch37_v4
MongoDB shell version: 3.0.9
connecting to: mongodb-dev/cellbase_hsapiens_grch37_v4clinical_variants
> show collections;
clinical_variants
conservation
gene
genome_info
genome_sequence
metadata
protein
protein_functional_prediction
protein_protein_interaction
regulatory_region
repeats
variation
variation_functional_score
```