Architecture

CellBase has been implemented in a very modular way providing a three layer architecture, enabling different access levels and use cases of the system:

1. MongoDB Databases in current implementation. An independent MongoDB database is created for each species and assembly. Collection schemas and indexes were designed to ensure an efficient response to the main expected queries (See MongoDB implementation for detailed specification of current mongo schema)
2. a Java abstraction layer that deals with the peculiarities of the underlying database system and provides an API to access the data (DB Adaptors). Advanced users may directly use this API to avoid network access and web services overload thus improving the performance of their implementations. Moreover, any advanced user wanting to use any other database system will just have to implement this API plugin for the desired database system. A cache system based on a Redis server is currently under development.
3. a set of REST Web Services and gRPC API which provide an efficient high level interface to the data regardless of the programming language. Most frequent access will be through this top layer. Thus, the Web Services API can be queried either by provided clients in Python, R, Java and JavaScript, the Java CLI or by own user-developed methods that implement calls to the Web Services by building appropriate urls. Comprehensive RESTful API specification is provided by Swagger: http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/ (Note: Python client is available in current develop branch, future 4.5.0 release. A first approach to the R client is also available and under development).

