PLCC Cluster Manual

About

This is a short manual for *PLCC cluster*, a collection of scripts which allows you to update the PTGL v3 server with fresh PDB data.

PLCC cluster runs the normal *plcc* java software for many PDB files on a computer cluster or on several machines in a LAN.

Versions

Available versions

Using PLCC cluster, you can run an update of the PTGL in different technical setups:

1. SINGLE: on a single machine (not recommended, most likely takes several weeks)
2. LAN: on some machines connected via LAN, which have a shared network drive (but no queue management software or any other cluster stuff installed)
3. CLUSTER: on a computer cluster which runs the openPBS queue. For this, two sub versions exits:
    1. CLUSTER_MPI: using openMPI. One job contains many (like 200) PDB files, and all cores of the node which is assigned the job work on them until all are done. The job is done when all PDB files are done.
    2. CLUSTER_NOMPI: without openMPI. Each job contains only a single PDB file, and MPI is not used.

Which version to use

Here is a rough comparison of the versions:

	System setup	Job management	Speed
SINGLE	Very easy	None	Very poor
LAN	Easy	None	Medium to OK
CLUSTER_MPI	Very complicated, MPI-specific tech problems	Full queue support, wall time, etc	Very good
CLUSTER_NOMPI	Complicated	Full queue support, wall time, etc	Very good

As of April 2015, we use the cluster version without openMPI (3.2 in the list above) to update the PTGL, and this is the recommended way to do it.

Running the Update

To run the update, you need to prepare some stuff first, independent of which method you intend to use. These preparations are describe in the following.

Common preparations

Before you start, ensure that PostgreSQL database server is installed correctly and running on the update machine or another server which is reachable from it. You should also create a database user and a database for update there (usually user *vplg*, database *vplg* owned by him). Then do the following:

- **Install PLCC cluster**: On the update machine (cluster head or whatever machine, usually NOT the PTGL server!), copy the plcc.jar file, splitpdb.jar file and the 'lib/' directory from your VPLG installation to the 'plcc_run' subdirectory. (You can also use subversion to get the latest version onto that machine.)
- **Install DSSP**: Install dsspcmbi on the machine, from the DSSP website
- **Get PDB data**: Get a copy of the PDB via rsync or update your local copy if required. You can use the 'update_files_via_rsync.sh' script for this.
- **Get DSSP data**: Get a copy of the DSSP via rsync or update your local copy if required. If you do not do this, the update procedure will have to run dsspcmbi for each PDB file to generate the DSSP data, which takes some additional time. It is recommended to download the data instead!
- **Update PLCC Cluster config**: Edit the config file 'settings_statistics.cfg' to match your local settings. Be sure to set the paths to dsspcmbi, and to the PDB and DSSP data. Also the plcc command line options should be chosen with care, especially output directory. Ensure it is reachable from all machines and writable for the user who runs the update.
- **Update PLCC config file**: Edit the PLCC config file (at '~/.plcc_settings', run 'plcc --help' once if you do not have that file yet) and configure the connection to the database server (machine, port, user, password).
- **Prepare database**: Change to plcc_run/ and execute './plcc NONE --recreate-tables' once to create the DB structure (and test DB connectivity).

Running the update using CLUSTER_NOMPI: