2019-08-15

Due to the number of people out on vacation, it was decided to defer the special topic discussion until later, and do a general discussion for this meeting.

Boulder

Mark M announced that we have hit a git-lfs limit for this monthly billing cycle (of which we have 23 days left). We currently have purchased 150 GB of storage and 150 GB of bandwidth (which means we can download up to 150 GB per month). We have hit the 150 GB bandwidth limit for this month. We will immediately purchase one more "data pack" for now to get our downloads working again. A "data pack" comes in 50 GB increments (both storage and bandwidth) and each data pack costs $5 per month.

Moving forward (in regard to the git-lfs limits), we need to clean up large data files. We are moving toward a testing scheme ("multi-tiered") that runs only the small, fast unit tests every time we clone and build something. The larger, more expensive tests will be run on a nightly or weekly basis. This testing scheme will help alleviate our demands on git-lfs since we can move large files associated with the nightly/weekly tests out of git-lfs (to S3, for example). Travis gave the example of SOCA which has high-resolution (1 degree grid) data in git-lfs which can be moved out and only used by the nightly/weekly tests. The multi-tier testing strategy will also help speed up download times since test data can be trimmed down.

One caution is that when you do move data out of git-lfs, you need to make sure those files are removed from the entire git history (and note that "git rm" is not sufficient to do this). Tools exist (bfg repo cleaner, for example) that help you wipe out the entire history of a file in git. More information and instructions will be provided.

Andy has successfully configured his laptop, and got ufo-bundle running (tests pass).

Travis is getting SOCA up and running on Discover for the purpose of running the larger test cases (eg, 1 degree grid).

Xin has a PR under review for the first stage (of 2 stages) of observation bias correction functionality.

Steve V has advanced the JEDI MPAS interface in regard to hydrometeors. He also has a functional C++ program that reads ODB files using the new ODC library from ECMWF. The first ODC task will be to compare performance with netcdf and the old ODB API library.

Steve H reported that Hamideh's update to the SST obs file has been merged in. This update reduces the number of obs from 200,000 to 200 for the purpose of speeding up unit testing.

Ming is developing a Fortran program to convert prepBUFR data into IODA netcdf. He is highly leveraging code from existing converters with the purpose of making his converter as generic as possible. If successful, this will eventually replace a number of converters with one executable that can be driven by a configuration file. The reader part of this code is working, and Ming is finishing up the netcdf writer section.

Maryam is building the capability to run automated testing with Amazon's CodeBuild services, as a potential alternative or complement to Travis-CI.

Mark O is making good progress on the JEDI workflow. His is writing this flow in Python 3, and is using GNSSRO COSMIC-2 data as a vehicle for development. Guillaume asked if Mark is coordinating with EMC (EMC has an existing workflow), and Mark responded that he will do so but he wants to wait until Yannick is back from vacation (next week) so that Yannick's project-level perspective can be included in the discussions.

EMC/GMAO

Guillaume report that SOCA is running inside EMC's workflow, and will coordinate with Mark O to converge the two workflow systems.

Ryan is preparing for next week's UFO/QC code sprint.

Hamideh reduced the obs file for SST as mentioned above.

Ben J initiated a discussion in regard to an issue reported by Jim R concerning bad data that got into the CRTM processing. It was decided to take this discussion offline. (After the meeting it was discovered that data marked as missing in the obs file got converted to JEDI internal missing data marks, which is what is supposed to happen, and a filter needs to be applied to remove these missing data before handed them off to CRTM.)

Jong has the TAU parallel profiler working in the container. He is still experiencing trouble on Theia due to some library incompatibilities. Mark M mentioned that the TAU profiler is from Oregon State, and has a freeware version. Jong is now implementing a packaged version of TAU in jedi-stack with Mark's help. It will soon be available for those who want to try it out. As a group we are still deciding on whether or not to purchase a license for a parallel debugger for use by the JEDI core team, the top choices being DDT and TotalView. If DDT is selected, it comes with a parallel profiler called MAP.

Met Office

Marek hasn't had a chance to investigate the Fortran style checker yet, but it's on his to-do list. Marek has a draft PR in UFO under review which contains time interpolation functionality.

Marek reported issues in Debug builds which are causing LFRic testing to fail. He suspects eckit 1.1.0 is involved, and added that OOPS, BUMP, UFO tests all pass (in Debug mode). It was noted that it's important to use both eckit (ECMWF version 1.1.0) and fckit (JCSDA fork, develop branch) together.

Mark M asked if it was okay to merge the config_mod PR in OOPS. Doing so means you have to use the fckit Fortran config interface (not the OOPS Fortran config interface, since this is removed in this PR). Marek said they have been holding off testing the PR until they get the issue mentioned above with LFRic resolved. Mark M said that we'll wait on the merge until we get the okay from the Met Office.

NRL

Sarah reported that they are debugging 3DVar increment errors, and they are helping Ben R get ready for next week's code sprint.