2020-01-09

Steve H opened the meeting by announcing that this meeting will be a general round-table update.

We began with Boulder. Mark M announced that JEDI applications should work on S4 again. They had been working previously but about a month ago something changed. First, the modules were not loading properly, which was causing build problems. This was fixed when the system administrators installed a newer version of lmod, at our request. The second problem has to do with parallel io with netcdf4/hdf5. The S4 sysadmins disabled file locking on the /data filesystem. Previously we had gotten around this by setting the HDF5_USE_FILE_LOCKING environment variable to FALSE but this was no longer working. As a result, the code would compile but many bump and fv3 tests would fail, if run from the /data directory. The /scratch filesystem did not have this problem - the tests passed when run from there. However, as of yesterday (Jan 8), the sysadmins enabled file locking on the /data filesystem as well. So, everything should work again as before, provided you specify srun as your mpi executable as described in the JEDI documentation. Steve H tried running fv3-bundle on the /data filesystem yesterday and confirmed that all tests pass. If you have any problems, contact Mark M.

Mark M also mentioned that there are new versions of both ecbuild and eckit coming, as will be described further by Steve H and Maryam. Steve H then elaborated on this idea by announcing that there is a pull request currently in ioda in which the test files are stored and accessed directly on the cloud (Amazon S3) instead of through git-lfs. The new version of ecbuild makes this access possible. So, after this pull request is merged, all repositories that use ioda data files (including ioda itself) will require the new ecbuild. The code may compile with older versions of ecbuild but many of the tests will likely fail.

In response to a question by Chris H, Mark M clarified what we mean by the **"new versions" of ecbuild and eckit**. We are now using the JCSDA forks for both of these repositories. If you are building eckit as part of your bundle and/or using an uninstalled version of ecbuild, then it is ok to just use the latest **jcsda::develop** branch of each. However, to track the versions of ecbuild and eckit that are installed in the containers and environment modules, we have begun using tagged versions of each. These tagged versions will generally end with a patch number of the form **.jcsda<x>**, where **<x>** is a number. For example, at the time of writing, the most recent tagged versions are **3.1.0.jcsda1 for ecbuild** and **1.4.0.jcsda2 for eckit** (**jcsda fork**). However, there are active pull request in each of these repositories so that may change in the coming week. If you intend to work with the tagged versions of ecbuild and eckit, navigate to the repository in question on jcsda GitHub and select the **releases** tab in the menu bar right above the code. Then choose the release with the most recent jcsda patch.

Maryam and Steve then elaborated on what this means for ioda. The new approach of storing ioda obs files on AWS S3 instead of git-lfs will require a new procedure for those who wish to add new test files. Maryam will write up instructions for this new procedure and we will soon publish them on the JEDI Documenation page. Stay tuned to these meetings and notes, as well at the JEDI GitHub Team page for updates.

Maryam went on to say that our changes to ecbuild now make it possible for us to store test files on any remote repository that can be accessed on the web (via curl or wget). We're using Amazon Web Services (AWS) S3 for now but this could be changed if needed. Maryam also mentioned that there is a new test that will check to see if the test data files are already available, and if not, it will download them from the remote repository. If this test determines that it must download and cannot do so, then the test will fail. Maryam has submitted pull requests for ufo, ioda and fv3 repos containing this new download test. If you are on a secure system that does not have access to external internet sites such as AWS S3, then you can upload the test files to your system separately.

Maryam added that there is a current pull request in saber that does the same thing, namely storing the test files on S3 instead of on git-lfs. But, most developers will be less impacted by this because the saber test files are only used by the saber tests but the ioda test files are used by multiple repositories. Still, after this PR is merged, many saber tests will fail if you are using an old version of ecbuild.

Maryam and Mark M then clarified what "access to AWS S3" means. The JEDI test files (beginning with ioda and saber but likely extended to other repos in the future) are stored in a public bucket on S3. So, you do not need an AWS account to access them - you only need internet access. And, if you are using up-to-date versions of ecbuild and ioda, this should be transparent - ecbuild will handle the file transfers for you automatically.

Xin then reported that his variable bias correction implementation for amsua and iasi is complete and he confirmed with Emily that it reproduces the GSI results. Next he plans to work on the pre-conditioning.

It was mentioned that the UKMO is using an older version of ecbuild so they will have to update to accommodate these changes. However, Mark M advised to wait a few days. The 3.1.0.jcsda1 version of ecbuild is not currently backward compatible with other ecmwf packages like odc. Maryam intends to fix that today. So, we plan to have a new **ecbuild** version **3.1.0.jcsda2** that should be available today or tomorrow. Once that is available, Mark M will generate new containers and environment modules and others are advised to do the same (or use the jedi-provided ones).

Chris H then asked if anyone had tried accessing AWS S3 from Hera. He had requested that the route to S3 be approved by NOAA but he had not tested it. Ryan confirmed that it is now working.

We then continued the round-table update in Boulder with Mark O. He has been making good progress with the new jedi-rapids workflow but has been forced to deal with issues regarding where to store the large volumes of data (much of it in fv3 background files). We had been storing these on AWS S3 but the associated data transfer costs (several thousand dollars for ~ 25 TB in December) might make this impractical, particularly considering that we are currently low on AWS credits. So, he has been developing a distributed file repository system that leverages the storage capacity now available on various HPC systems, in particular S4, Discover, and Cheyenne. Each model will have a preferred HPC platform for running the jedi-rapids workflow and the associated data files will be stored locally there. Mark has developed applications for monitoring, querying, synchronizing, and accessing the distributed data repositories. He expects to have the jedi-rapids workflow working soon on these three systems. In support of this effort, Dan has generated fv3 gfs background files at 3-hour intervals for July, 2019 and at 6-hour intervals from then until the present, with more on the way.

Steve Vahl has been working on the writer side of the ODC interface and has that almost done. This will allow him to enter a new phase of performance testing.

Chris Snyder and MMM colleagues have been working on integrating ABI observation files for use with mpas and jedi. Some issues that they are facing include where/how to handle super-obbing and cloud detection.

Travis is working on a poster for AMS.

Andy reported that he has a prototype geometry test working for land da.

We then moved on from Boulder to the joint EMC/GMAO gathering in Maryland. Dan has been working on developing a real-time cycling system for fv3jedi-geos. He's focusing on conventional obs for now and has made good progress. He was having convergence problems with gfs-16 and ABI data conversion. Some discussion followed with Chris S and Dan said he could make the ABI files available on S3 for the MPAS group to work with.

Ryan has been developing a new JEDI build system. When it is more mature (soon, he expects), he can give us an overview of what it is capable of and how to use it.

We then moved on to the UKMO. Steve S has continued working with the LFRic-jedi interface and plans to work more on developing the new UM-JEDI interface. He also announced a new member of the UKMO JEDI team, Kristen.

Wojciech has several active pull requests in ufo, including an extension of his recent development of thinning filters. He also mentioned that he has a new pull request on the JCSDA fork of eckit that he would like to be merged before we generate new containers and environment modules. Mark M agreed to take a look (other reviewers are asked to please take a look promptly so we can get to work generating the new containers - it's a minor change).

We then turned to NRL. Sarah said that they ran into problems building in release mode but Benjamin is helping them out with that. Sarah also announced a new team member, Eric Platt. Nancy reported that they have been making progress on integrating the NRL observational data stream into JEDI by writing out ioda data files. They expect to have a cycling 3DVar system running by the end of March.

Steve H asked if they plan to share the data conversion applications they develop and invited them to contribute to the ioda-converters repository. Nancy confirmed their intention to do so.

Sarah also added that the Neptune model team is working on refactoring the model and, as part of this effort, they plan to develop a version of the code that can be hosted on the JCSDA GitHub site. This will facilitate NRL's participation in the JEDI continuous integration testing and code review framework.

Chris H has made substantial progress on the shallow water model. The 3DVar and 4DVar applications are now mature and he has been using them to test the forecast accuracy of the model in various configurations at various resolutions, using a set of high-resolution simulation data as a true state. Everything seems to be working as expected, with 3DVar and 4DVar progressively improving the forecast accuracy. Next he plans to work on the B Matrix. Up to now he has been using an Identity matrix but he has begun implementing a Gaussian correlation model. He says the model, as written, poses some challenges for DA. For example, it has no external forcing so forecast errors are damped. Furthermore, the errors exhibit a periodicity that he attributes to inertial waves ("sloshing") in the domain. So, he is working on ways to address these challenges.

Then Steve H brought the meeting to a close. He announced that there will be **no meeting next week (Jan 16)** since many of us will be at the AMS meeting in Boston. This means that the Jan 23 meeting might be a good opportunity to do a focused topic discussion. Steve invited all JEDI meeting participants (including those not currently tuned in) to please let us know what topics you would like to discuss either at the Jan 23 meeting or in the future. And, Steve welcomed the new JEDI team members.

Best wishes to all for a peaceful and productive 2020!