

Running models on GFDL infrastructure: FRE 101

Raphael Dussin
MOM6 tutorial
October 17 2022



Let's download the examples

```
ssh gaea  
module load fre/bronx-19
```

```
git clone https://github.com/raphaeldussin/simple_xml  
cd simple_xml  
more README.md | grep fremake
```

```
>>> fremake -x ocean_only_experiments.xml -p ncrc4.intel18 -t prod MOM6_compile  
>>> fremake -x ice_ocean_experiments.xml -p ncrc4.intel18 -t prod MOM6_SIS2_compile
```

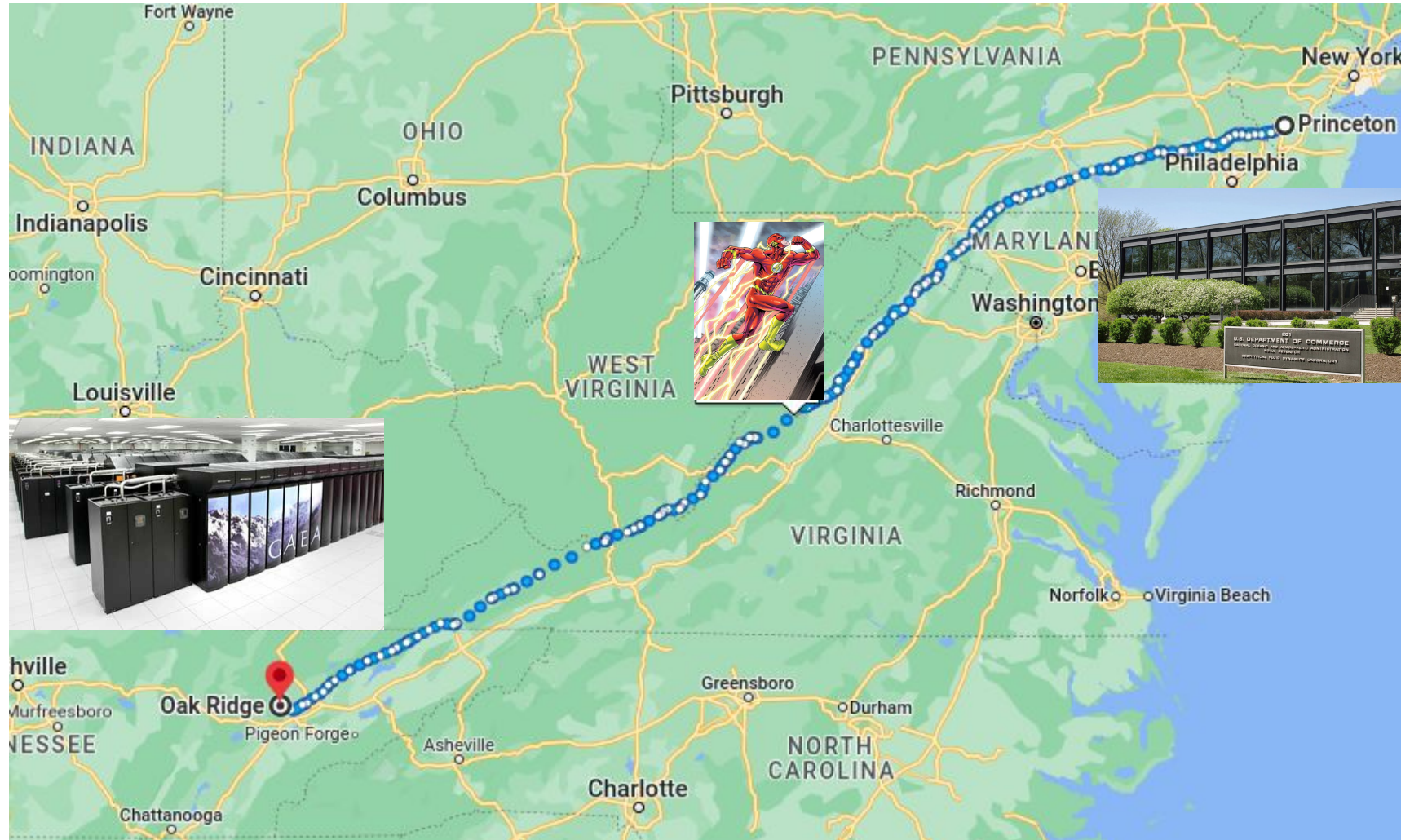
Overview of infrastructure

NCRC (gaea):

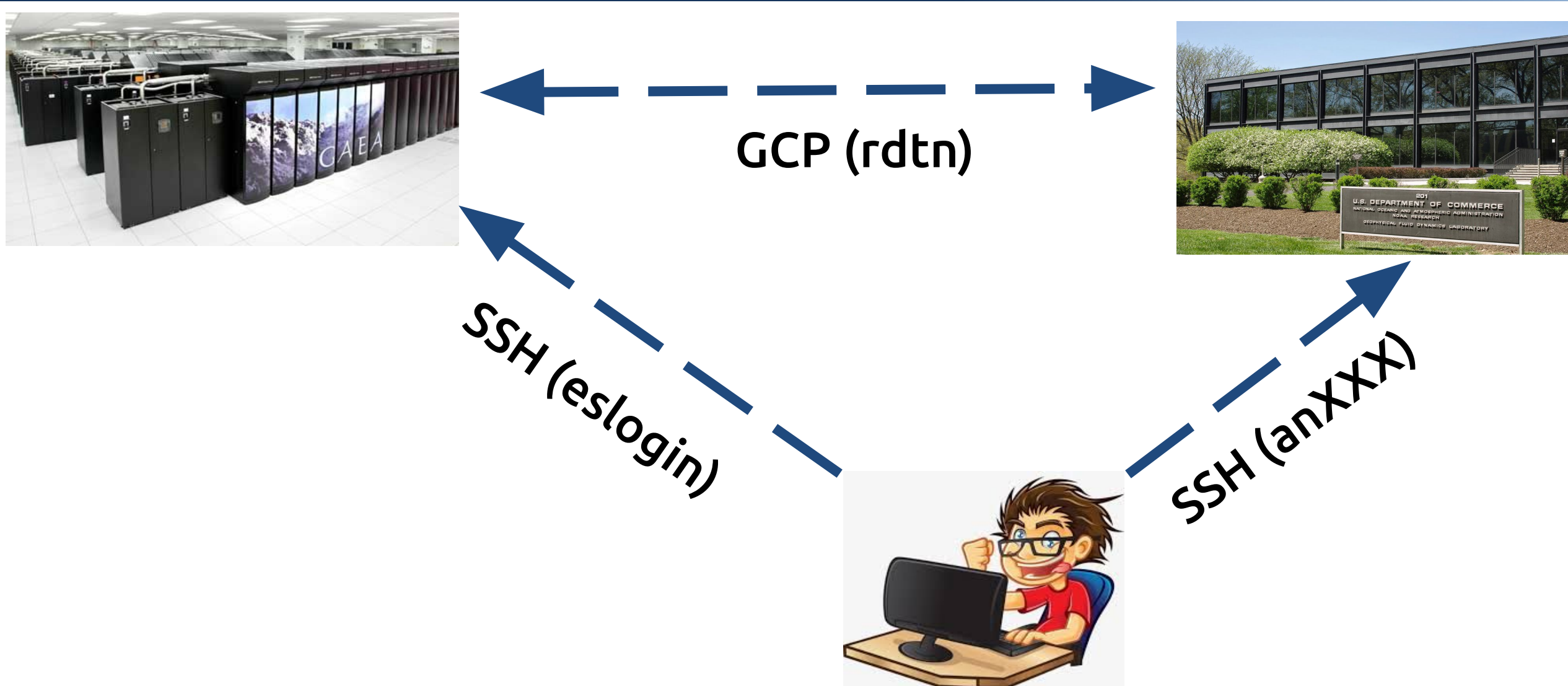
- 150k Intel cores
- c3, c4, c5 partitions
- /lustre/f2

GFDL (ppan/analysis):

- 100 nodes pp
- 30 nodes an
- /archive



Overview of infrastructure



Typical workflow for ocean/climate model



Compile the model **FREMAKE**

Run the model (batch) **FRERUN**

output.stager (rdtn/ldtn):

- recombine output (IO_layout)
- prepare restarts for next job
- create tar file for transfer
 - history
 - restart
 - ascii (logs)

Transfer to GFDL **/archive/initials** (rdtn)

creates **history**, **ascii**, **restart** directories

GCP (rdtn)



Run the refineDiags

- atmosphere: masking + derived quantities
- ocean: heat transports, overturning,...
- **gfdlvitals**: metrics (3d average temperature,...)

creates **history_refineDiag** directory and **db** (/home)

FREPP Create time series/averages for specified components (e.g. ocean_monthly, atmos_daily,...)

creates **pp** directory and its subdirectories

optionally, run analysis script and post figures in **/nbhome**

Naming conventions and directory structure

(sub)directories follow the same structure:

root/FMStag/Experiment/[gfdl.]platform-runtime/subdir

root: /archive/user , /home/user , /lustre/f2/dev/user, ...

FMStag: gives information on version of FMS and model

Experiment: self

platform: ncrc4 (c4), ncrc3 (c3),...

runtime: prod/repro [-openmp]

```
PPAN: Raphael.Dussin@an201 gfdl.ncrc4-intel18-prod-openmp: pwd
/archive/Raphael.Dussin/FMS2019.01.03_devgfdl_20210706/CM4_piControl_c192_OM4p125_v7/gfdl.ncrc4-intel18-prod-openmp
PPAN: Raphael.Dussin@an201 gfdl.ncrc4-intel18-prod-openmp: ls
ascii history history_refineDiag pp restart
```


What's where?



/lustre/f2/pdata/gfdl/gfdl_O/datasets : all model datasets/forcings

/lustre/f2/**dev**/user/FMStag

- model sources + compile script + executable
- experiment/platform-runtime
 - scripts/run : run script
 - state/run : state of model

/lustre/f2/**scratch**/user/**work**/FMStag

this is where the jobs are running, scrubbed

/lustre/f2/**scratch**/user/**FMStag**/experiment

- staging location for output/restarts (archive)
- scrubbed
- slurm logs

What's where?



/home/user/ncrc : copy of xmls

/home/user/FMStag/experiment:

- pp state, scripts and logs
- db for DORA

/nbhome/user/FMStag/experiment

- analysis figures
- copy of mom6-examples and other src (cvdp, mdtf,...)

/archive/user/FMStag/experiment

- history (tar)
- pp (time series/average netcdf files)
- restart (tar)
- ascii (tar)

Compiling and running the model with FRE

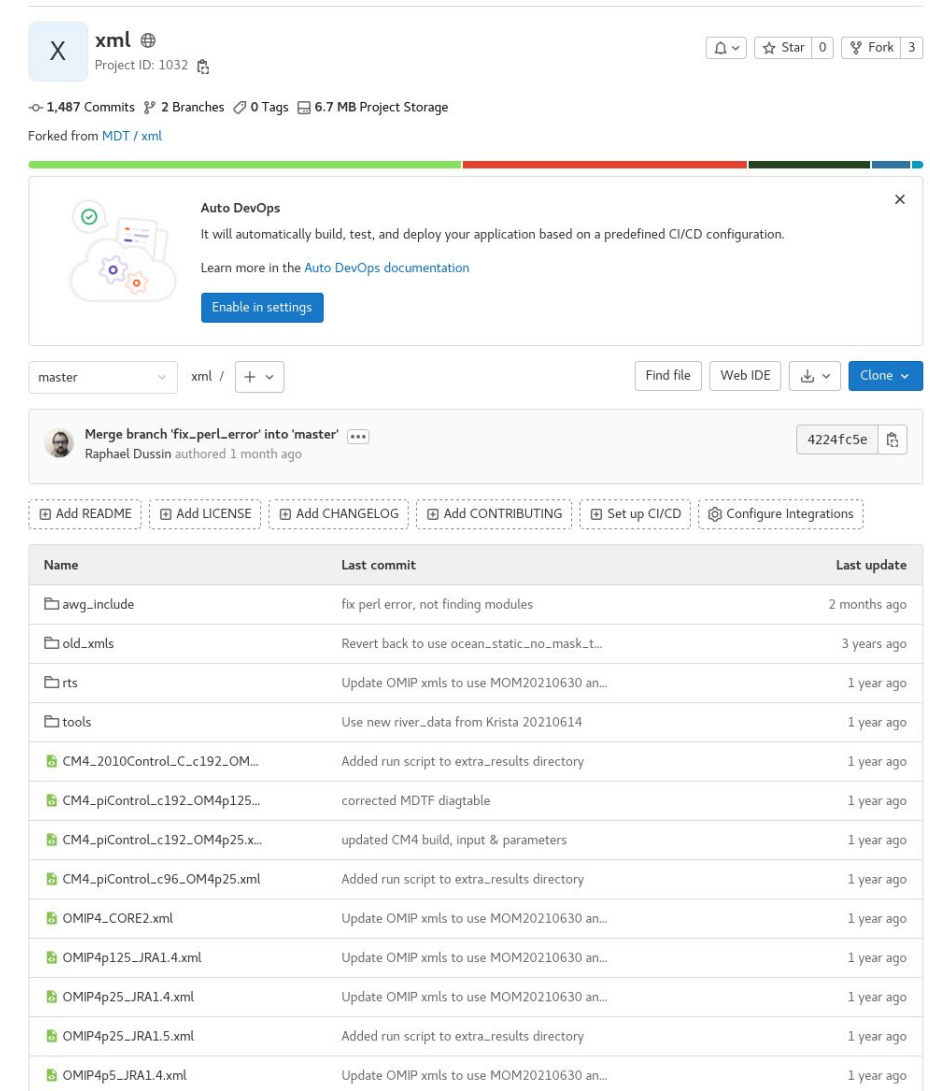
Everything is contained in the xml (and includes)!

- compilation instructions
- runtime parameters and overrides
- input files
- processor count and mask table
- diag_table
- post-processing instructions

1. Fork and clone xml repo (many forks):

`git@gitlab.gfdl.noaa.gov:ogrp/xml.git`

2. Modify from existing experiment



The screenshot shows the GitLab interface for the 'xml' repository. At the top, it displays the repository name 'xml' with a globe icon, project ID '1032', and statistics: 1,487 commits, 2 branches, 0 tags, and 6.7 MB project storage. It also shows 'Forked from MDT / xml'. Below this is an 'Auto DevOps' section with a green checkmark icon and text stating it will automatically build, test, and deploy the application. A blue button 'Enable in settings' is present. The repository is currently on the 'master' branch. A recent commit by Raphael Dussin is shown: 'Merge branch 'fix_perl_error' into 'master'' with commit hash '4224fc5e'. Below the commit list are buttons for 'Add README', 'Add LICENSE', 'Add CHANGELOG', 'Add CONTRIBUTING', 'Set up CI/CD', and 'Configure Integrations'. A table of files and their commit history is shown at the bottom.

Name	Last commit	Last update
awg_include	fix perl error, not finding modules	2 months ago
old_xmIs	Revert back to use ocean_static_no_mask_t...	3 years ago
rts	Update OMIP xmIs to use MOM20210630 an...	1 year ago
tools	Use new river_data from Krista 20210614	1 year ago
CM4_2010Control_C_c192_OM...	Added run script to extra_results directory	1 year ago
CM4_piControl_c192_OM4p125...	corrected MDTF diagtable	1 year ago
CM4_piControl_c192_OM4p25.x...	updated CM4 build, input & parameters	1 year ago
CM4_piControl_c96_OM4p25.xml	Added run script to extra_results directory	1 year ago
OMIP4_CORE2.xml	Update OMIP xmIs to use MOM20210630 an...	1 year ago
OMIP4p125_JRA1.4.xml	Update OMIP xmIs to use MOM20210630 an...	1 year ago
OMIP4p25_JRA1.4.xml	Update OMIP xmIs to use MOM20210630 an...	1 year ago
OMIP4p25_JRA1.5.xml	Added run script to extra_results directory	1 year ago
OMIP4p5_JRA1.4.xml	Update OMIP xmIs to use MOM20210630 an...	1 year ago

Step-by-step: compile

fremake download all code and prepare compile script

- -x xml file -t prod/repro [-openmp] -p ncrc[3,4].intel18 **compile**

```
(base) NCRC: Raphael.Dussin@gaea14 ~/RUNS/xml: module load fre/bronx-19
(base) NCRC: Raphael.Dussin@gaea14 ~/RUNS/xml: frelist -x OMIP4p25_JRA1.5_debug.xml
MOM6_SIS2_compile
OM4p25_JRA55do1.5_test1958_1 INHERITS FROM MOM6_SIS2_compile
OM4p25_JRA55do1.5_test1958_2 INHERITS FROM OM4p25_JRA55do1.5_test1958_1
(base) NCRC: Raphael.Dussin@gaea14 ~/RUNS/xml: fremake -t prod -p ncrc4.intel18 -x OMIP4p25_JRA1.5_debug.xml MOM6_SIS2_compile
Using source directory = /lustre/f2/dev/Raphael.Dussin/FMS2019.01.03_mom6_20201020//MOM6_SIS2_compile/src...
Cloning https://github.com/NOAA-GFDL/FMS.git on branch/tag 2019.01.03
Cloning https://github.com/NOAA-GFDL/ocean_BGC.git on branch/tag master
Note: switching to '199f2f24'.
```

compile script is under /lustre/f2/dev and then can be submitted:

```
Cloning https://github.com/NOAA-GFDL/atmos_null.git on branch/tag master
Cloning https://github.com/NOAA-GFDL/coupler.git on branch/tag 2020.03
TO SUBMIT => sleep 1; sbatch /lustre/f2/dev/Raphael.Dussin/FMS2019.01.03_mom6_20201020//MOM6_SIS2_compile/ncrc4.intel18-prod/exec/compile_MOM6_SIS2_compil
e.csh
(base) NCRC: Raphael.Dussin@gaea14 ~/RUNS/xml: █
```


frerun create run script:

- [-Z] -x xml file -t prod[-openmp] -p ncrc[3,4].intel18 **experiment**
- options: --overwrite, --extend

```
(base) NCRC: Raphael.Dussin@gaea14 ~/RUNS/xml: frelist -x CM4_piControl_c192_OM4p125_v7_init_tests_1mo.xml
cm4_sis2_compile
CM4_piControl_c192_OM4p125_v7_PHC_restart INHERITS FROM CM4_piControl_c192_OM4p125_v3
CM4_piControl_c192_OM4p125_v7_PHC INHERITS FROM CM4_piControl_c192_OM4p125_v3
CM4_piControl_c192_OM4p125_v7_noHenry INHERITS FROM CM4_piControl_c192_OM4p125_v3
CM4_piControl_c192_OM4p125_v4_hydrography20210614 INHERITS FROM CM4_piControl_c192_OM4p125_v3
CM4_piControl_c192_OM4p125_v4_protocol INHERITS FROM CM4_piControl_c192_OM4p125_v3
CM4_piControl_c192_OM4p125_v3 INHERITS FROM cm4_sis2_compile
(base) NCRC: Raphael.Dussin@gaea14 ~/RUNS/xml: frerun -x CM4_piControl_c192_OM4p125_v7_init_tests_1mo.xml -t prod-openmp -p ncrc4.intel18 CM4_piControl_c1
92_OM4p125_v7_PHC --overwrite
WARNING: Option --cluster=c4 has been automatically added
WARNING: Option --qos=normal has been automatically added
WARNING: Option --transfer has been automatically added
WARNING: The output staging type isn't defined - using the 'chained' one...
WARNING: The production run length (1 months) < 1 year - the postprocessor won't be called
WARNING: The 'make_exchange_reproduce' is .TRUE. in the 'xgrid_nml' namelist, which contradicts with absence of 'repro' in your targets
TO SUBMIT => sleep 1; sbatch /lustre/f2/dev/Raphael.Dussin/FMS2019.01.03_devgfdl_20210706/CM4_piControl_c192_OM4p125_v7_PHC/ncrc4.intel18-prod-openmp/scri
pts/run/CM4_piControl_c192_OM4p125_v7_PHC
(base) NCRC: Raphael.Dussin@gaea14 ~/RUNS/xml: █
```


Modifying your experiment: runtime parameters

- Simulation length **PROD_SIMTIME**
- For MOM/SIS: add #override param =
- For other component, edit namelist bloc. **CAREFUL: changing one entry resets all others to defaults**
- Restart controlled by \$restart_flag
 - n for first segment
 - r after
 - can be forced to n/r (**careful here too**)

```
<namelist name="MOM_input_nml" >
  output_directory = './',
  input_filename = '$restart_flag'
  restart_input_dir = 'INPUT/',
  restart_output_dir = 'RESTART/',
  parameter_filename = 'INPUT/MOM_input','INPUT/MOM_layout','INPUT/MOM_override'
</namelist>
<namelist name="SIS_input_nml" >
  output_directory = './',
  input_filename = 'r'
  restart_input_dir = 'INPUT/',
  restart_output_dir = 'RESTART/',
  parameter_filename = 'INPUT/SIS_input','INPUT/SIS_layout','INPUT/SIS_override'
</namelist>
<csh><![CDATA[
cat > $work/INPUT/MOM_override << MOM_OVERRIDE_EOF
! 1/4 degree parameter changes
#override DT THERM = 7200.0
#override TIDEAMP_FILE = "tidal_amplitude_TPX09_0M4p25.v20210125.nc"
#override KAPPA_H2_FACTOR = 0.93
#override CHL_FILE = "seawifs-clim-1997-2010.1440x1080.v20180328.nc"
#override CHL_VARNAME = "chlora"
#override U_TRUNC_FILE = "MOM_U_truncations"
#override V_TRUNC_FILE = "MOM_V_truncations"
#override USE_generic_tracer = False
#override USE_RIVER_HEAT_CONTENT = False
#override USE_CALVING_HEAT_CONTENT = False
#override ENERGYSAVEDAYS = 1
#override TOPO_FILE = "topog.nc"
!parameters for using /ncrc/home2/John.Krasting/diag_z3.nc
#override NUM_DIAG_COORDS = 3
#override DIAG_COORDS = "z Z ZSTAR", "rho2 RH02 RH0", "z3 Z3 ZSTAR"
#override DIAG_COORD_DEF RH02 = "FILE:diag_rho2.nc,interfaces=rho2"
DIAG_COORD_DEF_Z3 = "FILE:diag_z3.nc,interfaces=z_i"
! Kappa Shear
#override VERTEX_SHEAR = True
#override USE_RESTRICTIVE_TOLERANCE_CHECK = True
```

Modifying your experiment: adding/changing input

changing initial conditions:

```
<experiment name="OM4p25_JRA55do1.5_ePBLRL19_BBL01_VS_JHLtol_repro_test" inherit="OM4p25_JRA55do1.5_ePBLRL19_BBL01_VS_JHLtol">
  <input>
    <dataFile label="initCond" target="INPUT/" chksum="" size="" timestamp="">
      <dataSource platform="$(platform)">/lustre/f2/dev/Raphael.Dussin/archive/FMS2019.01.03_mom6_20201020/OM4p25_JRA55do1.5_ePBLRL19_BBL01_VS_JHLtol_repro_test.ncrc4-intel18-prod/restart/20180101.tar</dataSource>
    </dataFile>
    <!--There is MOM6/SIS2 restart provided in initCond.-->
    <namelist name="coupler_nml">
      months = $months,
      days   = $days,
      current_date = 2018,1,1,0,0,0,
      hours   = 0
    </namelist>
  </input>
</experiment>
```

adding more input files:

```
<dataFile label="input" target="INPUT/" chksum="" size="" timestamp="">
  <dataSource platform="$(platform)">/lustre/f2/dev/Raphael.Dussin/input/coords/diag_rho2_74layer.nc</dataSource>
</dataFile>
```

Quick note on inheritance

Experiments can inherit properties from previous experiment:

- At minimum inherit from "compile" experiment -> get executable
- Sensitivity can inherit from a reference simulation
- Repeated forcing: cycle2-N can inherit from cycle1

CAREFUL when using inheritance!

- checking the resulting script is a good idea
- breaking into a new xml at your discretion

Modifying your experiment: adding diagnostics

Option 1: add to existing/new diag_table

```
<dataFile label="diagTable" target="INPUT/" checksum="" size="" timestamp="">
  <dataSource platform="$(platform)">$(AWG_INPUT_HOME)/diag_table/diag_table_vortbudget</dataSource>
</dataFile>
<dataFile label="diagTable" target="INPUT/" checksum="" size="" timestamp="">
  <dataSource platform="$(platform)">$(AWG_INPUT_HOME)/diag_table/diag_table_MDTF</dataSource>
</dataFile>

<!-- CMIP6 atmos monthly fields -->
<dataFile label="diagTable" target="INPUT/" checksum="" size="" timestamp="">
  <dataSource platform="$(platform)">$(AWG_INPUT_HOME)/diag_table/cmip6/diag_table_Amon</dataSource>
</dataFile>
```

Option 2: append in xml

```
<diagTable order="append">
  "ocean_annual_rho2l35", 12, "months", 1, "days", "time"
  "ocean_model_rho2l35", "umo", "umo", "ocean_annual_rho2l35", "all", "mean", "none", 2
  "ocean_model_rho2l35", "vmo", "vmo", "ocean_annual_rho2l35", "all", "mean", "none", 2
</diagTable>
```

careful: job will die if duplicates in diag_table

Modifying your experiment: changing post-processing

Control what is included in the pp directory (time series/averages)

Can be in xml or include (look for postProcess)

Optionally generate regridded output and/or analysis

```
<!-- MONTHLY -->
<component type="ocean_monthly" start="$(PP_START_YEAR)" source="ocean_monthly" sourceGrid="ocean-tripolar" cmip="on">
  <timeSeries freq="monthly" chunkLength="$(CHUNK_LENGTH_A)">
    <!--analysis switch="$(ANALYSIS_SWITCH)" cumulative="no" script="$(AWG_INPUT_HOME)/analysis/stub/om4p125_monthly.frepp"/-->
  </timeSeries>
  <timeAverage source="monthly" interval="$(CHUNK_LENGTH_A)"/>
</component>
<component type="ocean_monthly_1x1deg" start="$(PP_START_YEAR)" source="ocean_monthly" sourceGrid="ocean-tripolar" xyInterp="180,360" cmip="on" >
  <dataFile label="xyInterpRegridFile">
    /archive/Raphael.Dussin/datasets/OM4p125/OM4p125_grid_20200921_noiceshelves_unpacked/remap_file_native_to_1x1.nc
  </dataFile>
  <timeSeries freq="monthly" chunkLength="$(CHUNK_LENGTH_A)" />
  <timeAverage source="monthly" interval="$(CHUNK_LENGTH_A)"/>
</component>
<component type="ocean_month_z" start="$(PP_START_YEAR)" source="ocean_month_z" sourceGrid="ocean-tripolar">
  <timeSeries freq="monthly" chunkLength="$(CHUNK_LENGTH_A)" source="ocean_month_z"/>
  <timeAverage source="monthly" interval="$(CHUNK_LENGTH_A)"/>
</component>
```


Modifying your experiment: changing the code

Compile "experiment" controls from where the source are pulled

Edit mom6 section
to add/checkout
your dev branch

May need to edit
paths and tags for
other components
(compatibility)

```
<experiment name="MOM6_SIS2_compile">
  <description>
    Make the executable for ocean-ice experiments.
  </description>
  <component name="fms" paths="FMS">
    <source versionControl="git" root="https://github.com/NOAA-GFDL">
      <codeBase version="$(FMS_TAG)"> FMS.git </codeBase>
      <csh><![CDATA[
        pushd FMS
        git checkout $(FMS_GIT_HASH)
        popd
      ]]></csh>
    </source>
    <compile>
      <cppDefs>-Duse_libMPI -Duse_netCDF $(F2003_FLAGS) -DMAXFIELDMETHODS_400</cppDefs>
    </compile>
  </component>

  <component name="mom6" requires="fms" paths="MOM6/config_src/{infra/FMS1,memory/dynamic_symmetric,ice/ice}">
    <description domainName="Ocean" communityName="GFDL-MOM6" communityVersion="dev/gfdl" communityUrl="https://github.com/NOAA-GFDL/MOM6">
    </description>
    <source versionControl="git" root="https://github.com/NOAA-GFDL">
      <codeBase version="dev/gfdl"> MOM6.git </codeBase>
      <csh><![CDATA[
        pushd MOM6
        git checkout $(MOM6_GIT_HASH)
        popd
      ]]></csh>
    </source>
    <compile>
      <cppDefs><![CDATA[ $(F2003_FLAGS) -D_MAX_FIELDS_100 -DNOT_SET_AFFINITY -D_USE_MOM6_DIAG -D_USE_MOM6_ICES -D_MOM6_GIT_VERSION_STRING=$(MOM6_GIT_HASH) ]]></cppDefs>
    </compile>
  </component>
```


Solving common problems: failed transfer to GFDL

Cause: network or bastion down/maintenance

Normal case: FRE recovers most times and resubmit transfers

Manual fix:

1. go to
 `/lustre/f2/dev/user/FMStag/experiment/platform-runttype/state/run`
2. remove any .lock files
3. module load fre
4. `output.retry $(pwd)`

Solving common problems: failed postprocessing

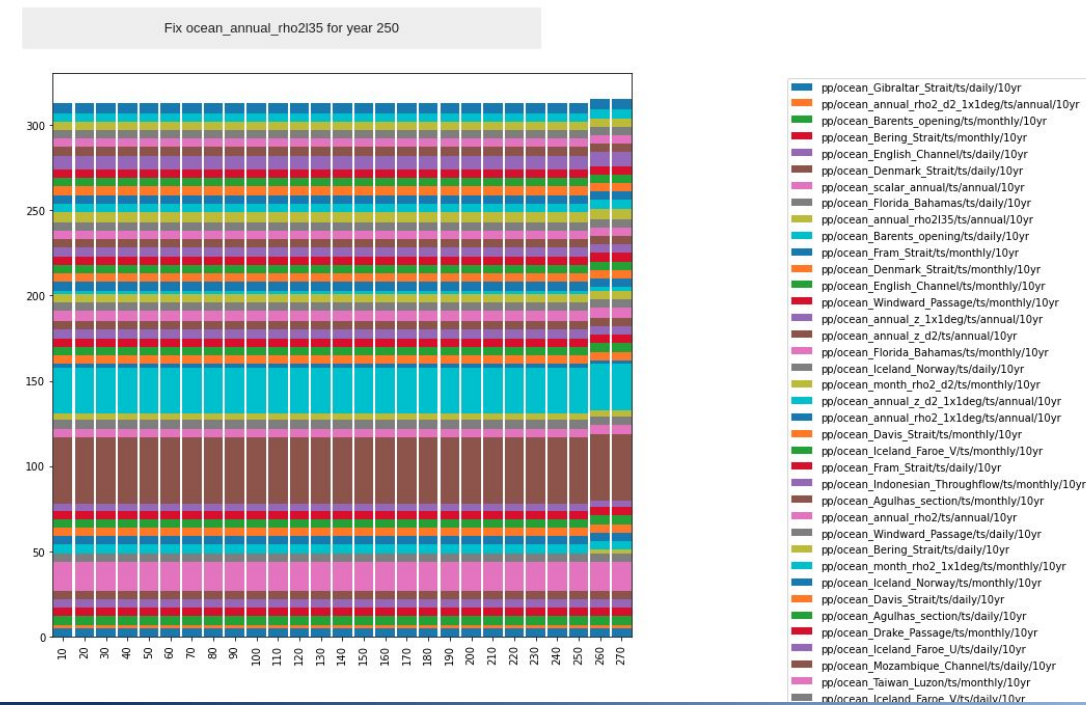
Most likely cause: refineDiag failed

Solution: in /nbhome/user/FMStag/experiment, replace mom6 by a recent clone of MOM6-examples and resubmit refineDiag script from /home/user/FMStag/experiment/platform-runttype/scripts/postProcess

Other issues: remove pp state file and resubmit individual scripts with **frepp**

```
frepp -s -p gfdl.ncrc4-intel18 -T prod-openmp -t 0001 --plus 9  
-x CM4_piControl_c192_OM4p125.xml  
-c ocean_monthly CM4_piControl_c192_OM4p125
```

Monitor/fix pp with dashboard template:
https://github.com/raphaeldussin/run_dashboard



Exercices with simple_xml

it's all in the README of simple_xml:
more README.md | grep frerun
for commands

- sensitivity: f-plane Phillips
- modify diag_table in Baltic
- add new vertical coord to Baltic

if you run out of time, go to the branch
with the solutions ;)

Exercices:

1. create a new Phillips experiment that **inherit** from the existing example, copy the **MOM_override** into the new of the xml and make it a f-plane experiment (i.e. **BETA** = 0.)
2. edit the Baltic example and replace the current diag_table by your own, in which you will save daily output for SST (tos), SSS (sos)
3. the Baltic sea has a maximum depth of 500 meters, finish the code snippet to build a custom **diagnostic** vertical coordinate with 5 meters resolution from the surface to 100 meters and 25 meters below 100 meters.

```
import numpy as np
import xarray as xr

interfaces =

thicknesses =

# define netcdf variables and write to file
vertcoord = xr.Dataset()

vertcoord["z_i"] = xr.DataArray(interfaces,
                                dims=("z_i"),
                                attrs={"long_name": "Interface target depth",
                                       "units": "m"})
vertcoord["dz"] = xr.DataArray(thicknesses,
                                dims=("z_i"),
                                attrs={"long_name": "z* coordinate level thickness",
                                       "units": "m"})

vertcoord.to_netcdf("diag_z_baltic.nc", encoding={"z_i": {"_FillValue": 1e+20},
                                                  "dz": {"_FillValue": 1e+20}})
```

Add this new coordinate to the **DIAG_COORDS** and add thetad and so to the diag_table.

advice: you can use my python install

```
alias ipython=/lustre/f2/dev/Raphael.Dussin/miniconda3/envs/repro/bin/ipython
```

if you're having issues creating the file (or we're running out of time), use the one I already created:

```
/lustre/f2/dev/Raphael.Dussin/input/MOM6tutorial/Baltic/diag_z_baltic.nc
```


May the FRE be with you!

Questions?



