Script for GraphCast demonstration (streamlined version)
ATM419/563 – Fovell -- Fall 2024

This demonstration involves the Google GraphCast machine learning weather prediction model, operated through ECMWF's ai-models front-end.

* ---------- **preliminaries** ---------- *
• make a new directory in your lab scape and move into it
$ mkdir GRAPHCAST
$ cd GRAPHCAST

• set up the environment
$ source /network/rit/lab/atm419lab/SOFTWARE/.bash_profile_2311
$ source /network/rit/lab/atm419lab/anaconda3/bin/activate
$ conda activate /network/rit/lab/atm419lab/anaconda3/envs/GCTEST

• copy and unpack the GRAPHCAST package

$ cp /network/rit/lab/atm419lab/GRAPHCAST/SETUP_GRAPHCAST.tar .
$ tar -xvf SETUP_GRAPHCAST.tar

• This creates the new directory **ai-models-graphcast**. Move into it
$ cd ai-models-graphcast

* ---------- **running the model** ---------- *
• you are in your $LAB/fall2024/**yournetid**/GRAPHCAST/ai-models-graphcast directory

• submit the batch job
$ sbatch -p burst-daes submit_graphcast

• outputs with names like "**out-XX.grib**" will be created. These are GRIB2 files. The out-0.grib file will appear first, and the rest of the files all at once when the model finishes

• Inside the submit_graphcast script, this line runs the model
ai-models --input file --file /network/rit/lab/atm419lab/GRAPHCAST/ERA5_2021123012.grib --date 20211230 --time 12 --path 'out-{step}.grib' --lead-time 18 graphcast

→ The file ERA5_2021123012.grib contains a subset of ERA5 reanalysis fields from 06 *and* 12 UTC 12/30/2021. GraphCast requires two initialization times 6 h apart
→ We specify the start time as 20211230 at 12 UTC
→ We specify the run time as 18 hours

• Model run takes about 10 min (absent resource competition)

• **tail -f sbatch.err.out** … will see some complaints like this. Ignore them

2024-04-24 10:49:30,062 INFO Unable to initialize backend 'cuda':
2024-04-24 10:49:30,062 INFO Unable to initialize backend 'rocm': module 'jaxlib.xla_extension' has no attribute 'GpuAllocatorConfig'
2024-04-24 10:49:30,063 INFO Unable to initialize backend 'tpu': INTERNAL: Failed to open libtpu.so: libtpu.so: cannot open shared object file: No such file or directory

• When you see something like the below, the model is wrapping up

2024-04-24 10:56:35,856 INFO Doing full rollout prediction in JAX: 7 minutes 5 seconds.
2024-04-24 10:56:35,856 INFO Converting output xarray to GRIB and saving

• At the end of the model run, these output files will have been created:
→ out-0.grib, out-6.grib, out-12.grib, out-18.grib

* ---------- **analyzing the output** ---------- *
• we will combine the model outputs into a single GRIB2 file "combined.grib2"

$ grib_copy out-0.grib out-6.grib out-12.grib out-18.grib combined.grib2

• copy the notebook
$ cp $LAB/GRAPHCAST/GRIB_plot_GC.ipynb .

• launch ARCC Jupyterlab. Batch or snow. Minimal resources suffice.
• execute notebook

• If you cannot run the model for some reason, you can edit the notebook to point to the pre-made output file:
/network/rit/lab/atm419lab/GRAPHCAST/combined.grib2

* ---------- **next steps (on your own)** ---------- *
(1) you need to establish an account on ECMWF CDS
https://cds.climate.copernicus.eu/how-to-api
(2) then you can run ai-models-graphcast fetching ERA5 data from CDS
→ this can be **very, very slow** to retrieve data from CDS
→ Edit submit_graphcast to alter the "ai-models" line
ai-models --input cds --date 20211230 --time 12 --path 'out-{step}.grib' --lead-time 18 graphcast &
(3) get it working on GPU instead so it runs very fast!!!

{Coureurs des GraphCast}