JASMIN Workshop: Exercise 04: A script to extract 1 variable from a file in the ERA-Interim dataset

Scenario

I am working on a project studying global cloud cover. I am looking at a case study and I have identified that the ERA-Interim data set is an appropriate source for my data. It is held on disk in the CEDA archive (located on JASMIN).

Objectives

Extract the total cloud cover ("TCC") variable for 1st January 2017 at midnight (00:00), from the ERA-Interim data set. The data files contain a large set of variables so I want to use the Climate Data Operators (CDO) tool in order to extract only the "TCC" variable.

JASMIN resources

- Scientific analysis server
- Space to store the output file: `/group_workspaces/jasmin2/workshop/users/$USER/ex04`
- Access to the CDO tool
- Read-access to the ERA-Interim data set in the CEDA archive - requires a CEDA account

Local resources

- SSH client (to login to JASMIN)

Instructions

1. Start ssh-agent session and add JASMIN private key
2. SSH to a scientific analysis server
3. Identify path to the required data file
4. Decide on the output file path
5. Activate the environment containing the CDO tool
6. Run the CDO tool to subset the file and extract the "`TCC`" variable
7. Quick-check the contents of the output file with "`ncdump`" tool

Review

This exercise demonstrates how to:

- Login to JASMIN and access the scientific analysis servers ("sci")
- Run a command-line tool interactively to read data from the CEDA archive
- Write outputs to a JASMIN Group Workspace

This is a basic workflow suitable for small tasks and setting up your processing. When the amount of processing increases then it makes good sense to move on to using the LOTUS batch cluster.

Alternative approaches and best practice

- Use the CEDA OpenDap server to extract the variable.
- Use other tools to run the extraction.
- Run as a batch job on LOTUS
- *Have any files been accidentally left on the system? (E.g. in* `/tmp/`*)*

Cheat sheet for Exercise 04: A script to extract 1 variable from a file in the ERA-Interim dataset

1. Start ssh-agent session and add JASMIN private key (skip if already done)

```
eval $(ssh-agent -s)
ssh-add ~/.ssh/id_rsa_jasmin
```

2. SSH to a scientific analysis server

```
ssh -A <username>@jasmin-login1.ceda.ac.uk
ssh jasmin-sci5 # Could use sci[123456]
```

3. Identify path to the required data file
    a. ERA-Interim surface analyses live under:
    `/badc/ecmwf-era-interim/data/gg/as/`
    b. Sub-directories under that are: `"YYYY/MM/DD/"`, locate file for 1st January 2017 at midnight (00:00).
    c. Set the input file:

```
INPUT_FILE=/badc/ecmwf-era-interim/data/gg/as/2017/01/01/ggas201701010000.nc
```

4. Decide on the output file path:

```
OUTPUT_FILE=/group_workspaces/jasmin2/workshop/users/$USER/ex04/output.nc
mkdir -p /group_workspaces/jasmin2/workshop/users/$USER/ex04
```

5. Activate the environment containing the CDO tool

```
module load jaspy
```

6. Run the CDO tool to subset the file and extract the "`TCC`" variable
    a. Consult the CDO manual to see how to extract a variable by name:
    https://code.mpimet.mpg.de/projects/cdo/embedded/index.html#x1-1460002.3.3

    b. Run CDO:

```
cdo selname,TCC $INPUT_FILE $OUTPUT_FILE
```

7. Quick-check the contents of the output file with "`ncdump`" tool

```
$ ncdump -h $OUTPUT_FILE

netcdf output {
dimensions:
        longitude = 512 ;
        latitude = 256 ;
        surface = 1 ;
        t = UNLIMITED ; // (1 currently)
variables:
        float TCC(t, surface, latitude, longitude) ;
                TCC:standard_name = "cloud_area_fraction" ;
                TCC:long_name = "Total cloud cover" ;
...
```

Alternative approaches and best practice

- Use the CEDA OpenDap server to extract the variable
    - You can access and subset data using the OpenDap protocol and the CEDA THREDDS server:
        - Using the web-GUI: http://dap.ceda.ac.uk/thredds/dodsC/badc/ecmwf-era-interim/data/gg/as/2017/01/01//ggas201701010000.nc.html
        - Scripting access to the required subset, see the Help pages: https://help.ceda.ac.uk/article/4442-ceda-opendap-scripted-interactions https://help.ceda.ac.uk/article/4712-reading-netcdf-with-python-opendap

- Use other tools to run the extraction
    - Many tools/libraries exist on JASMIN, such as:
        - python-netCDF4
        - cf-python
        - iris
        - nco
        - R
        - IDL
    - See more information about software on JASMIN at: https://help.jasmin.ac.uk/article/273-software-on-jasmin

- Run as a batch job on LOTUS:
    - See the exercise 5 to find out how.

- *Have any files been accidentally left on the system? (E.g. in* `/tmp/`*)*
    - It is important to clean up any temporary files that you no longer need.
    - Please check whether the tools you use have left any files in "`/tmp/`".