EIS SMCE Data Conversion Software Task Template

Version: 0.3

Overview

The EIS SMCE Data Converter is a python application implementing a parallel workflow for converting EIS datasets on adapt/discover to zarr format.

Dependencies:

EIS SMCE depends on the following python packages which are all available from pip or *conda_forge*:

awscli	**boto3**	**dask**	**intake**	
intake-xarray	**numpy**	**pyhdf**	**rasterio**	**dask-jobqueue**
rioxarray	**s3fs**	**traitlets**	**xarray**	**zarr**

EIS SMCE Package Installation and Update

Install eis_smce as follows:

```
>> cd <install_dir>
>> git clone https://github.com/nasa-nccs-cds/eis_smce.git
>> cd eis_smce
>> python setup.py install
```

The installation process places sample copies of all eis_smce run configuration files **(*.cfg)** in the directory **~/.eis_smce/config**. This should work fine for testing, but in practice all run config files should be located outside of the container (so edits can be preserved).

Update and Re-install eis_smce as follows:

```
>> cd <install_dir>/eis_smce
>> git pull
>> python setup.py install
```

The update and re-install process should be performed each time the container is (re)built.

Running EIS_SMCE on discover

Before executing the container the user should start a slurm session:

```
>> salloc --nodes=1 --constraint="sky|hasw" --time=12:00:00
```

Inside the slurm session the user should use the container to execute an equivalent of the following:

```
>> conda activate eis_smce       # ignore this line if using pip
>> cd <install_dir>/eis_smce
>> python ./workflows/zarr_conversion.py ~/.eis_smce/config/zc.cfg
```

The zarr_conversion script takes (as an argument) the path to an eis_smce config file named (by default) zc.cfg. The eis_smce installation process installs a sample version of this file which should be copied/ edited to set the parameters (e.g. input/output paths) for each dataset that will be processed by the workflow.

Repository

https://github.com/nasa-nccs-cds/eis_smce

Branch Name

main

Software Design

Extends the *Intake-xarray* framework.

Tests

A test workflow can be run (within a slum session) on discover by executing:

```
>> cd <install_dir>/eis_smce
>> python ./workflows/zarr_conversion.py ~/.eis_smce/config/zc.cfg
```